Integrating mathematical models with experimental data from *in vivo* studies to investigate the within-host dynamics of *Salmonella* Typhimurium

Myrto Vlazaki
Department of Veterinary Medicine
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy
in Mathematical Biology

Newnham College 2020

I would like to dedicate this thesis to my loving mum and grandma

Declaration of length and originality

This dissertation is the result of my own work and includes nothing which is the outcome of work done by others except where specifically indicated in the text. This work has not been submitted in whole or in part for consideration for any degree, diploma or any other qualification at this, or any other university. This thesis does not exceed the word limit of 60,000 words excluding figures, tables, appendices and bibliography.

Myrto Vlazaki
2020

Thesis Summary/Abstract

Author's name: Myrto Vlazaki

Thesis Title: Integrating mathematical models with experimental data from *in vivo* studies to investigate the within-host dynamics of *Salmonella* Typhimurium

Bacterial diseases have historically accounted for a high burden of morbidity and mortality worldwide. While antibiotic therapy and immunisations have significantly improved clinical outcomes in the last couple of centuries, the rate at which new treatments and vaccines are being developed has slowed down. One way to accelerate innovation in the field is by understanding the host-bacteria interactions that take place *in vivo* and affect the clinical outcome of the disease, the response to treatment and the immunological response following vaccination. Mechanistic mathematical models are being increasingly used to advance our understanding about unobservable, host-pathogen interactions *in vivo*. These models allow us to develop hypotheses about the underlying biological mechanisms responsible for an observable and measurable outcome (e.g. the number of bacteria in a tissue of interest) and use them to try to match what has been experimentally or clinically observed.

In this thesis, I have developed mechanistic models and used them in conjunction with experimental data on systemic murine *Salmonella* Typhimurium infections to:

1. characterise the effects of different vaccine preparations on the immune system of the host and its response to a secondary bacterial challenge,

2. compare the effects of different antibiotic classes on bacterial populations residing in different organs, and predict whether the residual bacteria following antibiotic treatment differ in their proliferative capacity,

3. develop a new, flexible methodological tool for *a priori* experimental design in studies of bacterial dynamics at the level of the population within the host.

Acknowledgements

I would like to wholeheartedly thank my supervisor, Dr Olivier Restif, for believing in me, developing my skills and building my confidence to work in an interdisciplinary field. Throughout these years, he has encouraged me to work methodically and independently, to question my methods and results, and to always explore new possibilities. His support has been ever-present and at the end of this endeavour I feel not only better equipped academically, but also more mature, resilient and confident on a personal level. Equally, I am grateful for the fruitful conversations and mentorship by Dr David Price, who was always someone to count on for guidance and words of encouragement!

Furthermore, I owe immense gratitude to my parents, Maria and George, who have supported me throughout my life, taught me to always seek new challenges, to embrace failure as an opportunity to learn, and to believe in making the impossible possible. Thank you, mum, for always reminding me to be honest with myself, to acknowledge my mistakes, to reflect, to change and to find solutions. A big thank you goes to my brother, Alex, who set high standards, paved the way, let me follow in his footsteps and always looked out for me!

I would also like to express my gratitude to Daniel, my closest friend, whose support and sense of humour made me look forward to a new day in the Department and the gym! Our daily scientific and non-scientific conversations kept me sane during the early months of lockdown and gave the motivation to carry on.

Finally, thank you to my partner, Ross, whose love and support made the challenging days seem easy and whose encouragement served as a never-ending source of motivation to set new goals and never stop trying!

As I am now back to medical school, I feel grateful to have been given the opportunity to expand my knowledge and skillset, and to work with such a supportive and enthusiastic team of individuals! Hoping to combine clinical work with research, I will be on the lookout for opportunities for collaborative, interdisciplinary research, and try to always encourage my colleagues to step out of their comfort zone and grow as people, clinicians and academics!

Table of Contents

i. Thesis Abstract

Bacterial diseases still remain **a major threat**, especially for vulnerable demographic groups in less privileged parts of the world. While bacterial infections are -in principle- treatable with antibiotics, they are becoming increasingly challenging to manage due to the rise of antibiotic resistance and tolerance to most available first-line antibiotics. To counteract this challenge, the global community is exploring ways to better **understand the host-pathogen interactions** underpinning pathogenesis, to **optimise the use of existing antibiotics** and to **accelerate the development of novel therapeutics and vaccines**.

Historically, much of microbiological research has been centred around gaining knowledge about the components involved in **host immunity** and **bacterial virulence**. The conventional approaches used by microbiologists, such as **genetic engineering**, **microscopy** and **culture techniques**, have allowed them to obtain static snapshots of bacterial populations or single bacterial cells. By combining qualitative data with **statistical analysis**, microbiologists can **establish correlations** between interventions and outcomes, but not infer the biological mechanisms responsible for the observed relations. For example, using statistical tests combined with snapshot data on the number of bacteria before and after antibiotic administration, it is possible to determine whether an antibiotic of interest is significantly correlated with lower bacterial numbers in a particular tissue, but it is not possible to gain further insights as to whether the observed reduction in bacterial loads is due to enhanced bactericidal activity, induction of bacteriostatic mechanisms or a combination thereof.

Moving a step further from the mere establishment of statistical significance, it is becoming increasingly feasible to combine microbiological data with more sophisticated, **mechanistic models** that can **address questions regarding the dynamical changes** underlying the snapshot data obtained through conventional observational techniques. Characteristic examples of such questions include:

1) What is the relative contribution of bacterial death and division events in shaping bacterial numbers in *in vivo* experimental observations?

2) When and how frequently do bacteria move between different tissues?

3) What is the directionality of inter-tissue colonisation in an organism, e.g. given the presence of bacteria in organs A and B, is it possible to deduce whether bacteria from tissue A have seeded tissue B, or vice versa?

Obtaining answers to these questions provides a more **complete understanding** of the within-host dynamics of bacterial infections in the naïve host, as well as in response to interventions; prophylactic or therapeutic. By **integrating experimental data** with **mathematical modelling,** it is possible to **maximise the biological insight** offered by experimental observations, without the need to expend more resources.

The present thesis uses experimental data on systemic infections with *Salmonella enterica* serovar Typhimurium to showcase the benefits of integrating *in vivo* microbiological data with mechanistic mathematical models to:

1) obtain complementary biological insights into the within-host dynamics of *S.* Typhimurium in the naïve, immunised and antibiotic-treated murine host, and

2) optimise resource allocation during experimental design in order to obtain maximally informative experimental data.

In particular, the aims of the present thesis can be summarised as follows:

Bacterial dynamics in the naïve host (Chapter 3):

- To formulate and compare a range of models corresponding to biological hypotheses regarding the **early dynamics** (0-6 hours) of systemic *S.* Typhimurium infections, especially with regards to early inter-tissue migration events and bactericidal activity in the blood.

- To quantify the **late-phase dynamics** (6-72 hours), which have hitherto remained quantitatively uncharacterised in the naïve host, due to the lack of adequate mathematical tools.

Bacterial dynamics in the immunised host (Chapter 4):

- To quantitatively **compare the effect of a live-attenuated and a whole-cell killed vaccine** on the within-host dynamics of *S.* Typhimurium systemic infections over a period of **72 hours**.

Bacterial dynamics in the antibiotic-treated host (Chapter 5):

- To quantitatively **compare the effect of a fluoroquinolone (ciprofloxacin) and a beta-lactam (ampicillin) antibiotic** on the within-host dynamics of *S.* Typhimurium systemic infections **during treatment** (time period of 3 days) and following **antibiotic withdrawal** (time period of 2 days).

- To determine **whether, when and where antibiotic-tolerant bacterial subpopulations arise** during antibiotic treatment.

- To characterise the **dynamics** of bacteria originating **from previously tolerant populations and antibiotic-sensitive populations** following antibiotic withdrawal.

Development of a technique to optimise experimental design (Chapter 6):

- To **develop** a **computationally efficient tool** to **optimise experimental design** with regards to **inference precision** using microbiological data from studies with isogenic tagged strains.

- To appropriately **test the performance** of the novel **tool**.

- To **showcase the use of this tool** in a range of biological scenarios of **exponential growth, decay and "source-to-sink" bacterial migration**.

- To illustrate the use of the tool in **optimising resource allocation** (e.g. number of animal hosts, number and choice of sampling time points, inoculum size) to **achieve maximal inference precision**.

ii. List of Tables

1.

3.

4.

5.

6.

iii. List of Figures

iv. List of Publications and Conference Presentations

Publications

1. **Vlazaki M**, Huber J, Restif O. Integrating mathematical models with experimental data to investigate the within-host dynamics of bacterial infections. Pathog Dis. 2019;77(8): ftaa001.

   Authors contribution
   M.V. prepared and drafted the manuscript. All authors critically appraised the review.

2. Rossi, O., **Vlazaki, M.**, Kanvatirth, P., Restif, O., & Mastroeni, P. Within-host spatiotemporal dynamic of systemic salmonellosis: Ways to track infection, reaction to vaccination and antimicrobial treatment. J of Micr Meth 2020; 176, 106008.

   Authors' contribution
   All authors contributed to conceptualization, writing - original draft preparation, writing - review and editing.

3. **Vlazaki M**, Rossi O, Price DJ, McLean C, Grant AJ, Mastroeni P, Restif O. 2020 A data-based mathematical modelling study to quantify the effects of ciprofloxacin and ampicillin on the within-host dynamics of Salmonella enterica during treatment and relapse. J. R. Soc. Interface 20200299.

   Authors' contributions
   M.V. conceived and performed the study, and drafted the manuscript. Ol.R. provided guidance throughout the study and helped with the writing process. Om.R. helped with data acquisition and organization. C.M., M.V. and D.J.P. developed the computational tools. P.M. and A.J.G. critically reviewed the manuscript.

4. **Vlazaki M**, Price DJ, Restif O. 2020 An experimental design tool to optimise inference precision in data-driven mathematical models of bacterial infections in vivo. Journal of The Royal Society Interface, 17(173), 20200717.

   Authors' contributions
   M.V. conceived and performed the study, and drafted the manuscript. Ol.R. and D.J.P. provided guidance throughout the study and helped with the writing process.

Conference Presentations

	Conference Title	Location	Date	Format	Title
1.	Cambridge Infectious Diseases Annual Meeting of Minds	Cambridge, U.K.	10/2018	Poster	Mathematical modelling to characterise the *in vivo* dynamics of *Salmonella* in the naïve host
2.	Annual Microbiology Society Conference	Belfast, U.K.	04/2019	Oral Presentation	Mathematical modelling to characterise the *in vivo* dynamics of *Salmonella* in the naïve and immunised host
3.	Cambridge Infectious Diseases Annual Meeting of Minds	Cambridge, U.K.	10/2019	Poster	Detection of antibiotic persistence from bacterial population-level data in a murine model of systemic salmonellosis
4.	EPIDEMICS Conference	Charleston, U.S.A.	12/2019	Poster	Mathematical modelling as a tool to optimise experimental design in within-host bacterial population studies using isogenic tagged strains
5.	Cambridge University Clinical Research Society Annual Conference	Cambridge, U.K.	04/2020	Oral Presentation	A quantitative comparison of the effects of beta-lactam and fluoroquinolone antibiotics on the within-host dynamics of *Salmonella* Typhimurium

aBC	approximate Bayesian Computation
C	Cytosine (DNA base)
CDC	Centre for Disease Control and Prevention
CFU	Colony Forming Units
cLN	caecal Lymph Node
DNA	DeoxyriboNucleic Acid
DNL	Draining Lymph Node
E. coli	*Eschericia coli*
FIM	Fisher Information Matrix
G	Guanine (DNA base)
HIV	Human Immunodeficiency Virus
HPC	Hepatitis C
Ig	Immunoglobulin
IL-6	Interleukin 6
iNTS	invasive Non-Typhoidal Salmonellosis
ITS	Isogenic Tagged Strain
KV	Killed Vaccine
LV	Live Vaccine
MIC	Minimum Inhibitory Concentration
MDE	Minimum Divergence Estimation
MLE	Maximum Likelihood Estimation
MLN	Mesenteric Lymph Node
NGS	Next Generation Sequencing
qPCR	quantitative Polymerase Chain Reaction
ROS	Reactive Oxygen Species
S. Typhimurium/*S.* Tm	*Salmonella enterica* serovar Typhimurium
TNF-α	Tumor Necrosis Factor α

1.
Background

Summary Points:

- Bacterial diseases are still major "killers" worldwide.
- Antibiotic resistance and tolerance render antibiotics less and less effective at definitively treating bacterial infections.
- Accelerated research and development of novel prophylactic and therapeutic agents is urgently needed.
- Deciphering the within-host dynamics of bacteria can help create targeted therapies and vaccines.
- Coupling experimental techniques with mechanistic mathematical modelling can maximise insights into the unobservable biological processes that collectively determine the fate of a bacterial infection.
- Computationally efficient inference tools are needed to enable the use of mathematical modelling in complex biological systems.
- Murine infections with *Salmonella* Typhimurium are a well-established experimental system used as an archetype to study the within-host dynamics of bacterial infections.

Relevant published output:

1. Vlazaki M, Huber J, Restif O. Integrating mathematical models with experimental data to investigate the within-host dynamics of bacterial infections. Pathog Dis. 2019;77(8): ftaa001. doi:10.1093/femspd/ftaa001

2. Rossi, O., Vlazaki, M., Kanvatirth, P., Restif, O., & Mastroeni, P. Within-host spatiotemporal dynamic of systemic salmonellosis: Ways to track infection, reaction to vaccination and antimicrobial treatment. J of Micr Meth 2020; 176, 106008. doi: 10.1016/j.mimet.2020.106008.

Declaration of Originality

This chapter includes written material that verbatim appears in a published review (Vlazaki, Huber and Restif, 2019), for which I certify that I am the sole contributor in terms of manuscript writing.

1.1. Bacterial Infections: Current Burden, Challenges and Prospects

1.1.1. Epidemiological Background

According to the most recent estimates by the World Health Organisation, bacterial diseases, including respiratory infections and diarrhoeal diseases, are classified amongst the top 10 leading causes of mortality, accounting for almost 6 million deaths globally every year worldwide (WHO, 2018) The burden of bacterial diseases is asymmetrically distributed in terms of geography, age, socioeconomic and health status, mostly afflicting infants, elderly and immunocompromised individuals in low-resource settings in Sub-Saharan Africa and South-East Asia (Yoshikawa, 2002).

While, in principle, most bacterial diseases can be either treatable with antibiotics, preventable with immunisation or both, in practice, the efficacy of antibiotic treatment and the feasibility of *en mass* vaccination is hampered by multifactorial challenges on a logistical and clinical level alike. Far from complacency, a successful response to bacterial diseases would ideally include rational management based on a mechanistic understanding of their progression *in vivo*.

1.1.2. Antimicrobial Therapy: Challenges to a Success Story

Antimicrobial therapy has been one of the most successful examples of chemotherapy in the history of humankind. Although commonly associated with the 20th century discoveries of Paul Erlich (Ehrlich and Hata, 1910) and Alexander Fleming (Fleming, 1929), concentrated traces of tetracycline in skeletons of prehistoric populations provide evidence that antibiotic agents have empirically been used long before the golden era of antibiotic revolution (Nelson *et al.*, 2010). However, despite the earlier intuitive use of naturally-occurring compounds with low level of antimicrobial activity, it was not until the systematic extraction and *de novo* synthesis of antibiotics in the 20th century that communicable diseases ceased being the leading cause of death globally (CDC, 2016).

Despite the plethoric discovery of antimicrobial compounds in the 20th century, the management of bacterial infections still remains precarious, largely due to the combined effects of antibiotic resistance (reviewed by Hofer, 2018) and tolerance (reviewed by Brauner *et al.*, 2016), compounded by unresolved inequities in antimicrobial access across the globe (The Center for Disease Dynamics, Economics and Policy, 2019). While antibiotic resistance and tolerance are different biological phenomena, they are both

responsible for bacterial unresponsiveness to antibiotic therapy. A clonal population of bacteria can be wholly insensitive to one or more antibiotics; alternatively, it can be sensitive albeit incompletely eradicated, leaving a surviving subpopulation capable of establishing chronic infections with future recrudescent bouts. The former phenomenon is termed *resistance* and is attributable to heritable genetic mutations enabling bacteria to circumvent antimicrobial action, while the latter is known as *tolerance* and encompasses the survival of non-genetic variants in a clonal bacterial population due to -what is currently seen as- non-heritable metabolic adaptations that are non-specific to the antibiotic (Levin-Reisman *et al*., 2019; Van den Bergh, Fauvart and Michiels, 2017).

To continue treating bacterial diseases effectively in the face of these challenges would require expedited drug research and development combined with stewardship of existing antibiotics (Doron and Davidson, 2011). Salient examples include efforts to optimise antibiotic administration regimens in terms of antibiotic class, timeframe and dose, and to maximise the synergistic effects between compounds in multi-therapy treatments, in alignment with endogenous immune control of the infection by the host. As the discovery and commercialisation of novel antibiotic agents is slow (Daron and Davidson, 2011; Silver, 2011), optimising the use of already available agents is an important mid-term solution. To achieve the synergistic orchestration of antibiotic activity and host immunity, there is a requirement for an integrated understanding of the dysfunctional host-pathogen interactions targeted by antimicrobial agents (Munguia and Nizet, 2017). Furthermore, there is a need to disentangle the complex interactions between host immunity and pathogen behaviour (Gjini and Brito, 2016), and to identify the determinants of successful disease establishment and progression (Munguia and Nizet, 2017; Casadevall and Pirofski, 2001).

1.1.3. Immunisation: Designing and delivering the "ideal" vaccine

Complementary with efforts to optimise antibiotic development and use, immunisation of vulnerable populations (Breiman *et al.*, 2012) has emerged as an equally promising population-wide strategy which has been used to control the spread and sequelae of bacterial diseases caused by important human pathogens such as *Mycobacterium tuberculosis, Haemophilus influenzae B, Streptococcus pneumoniae, Clostridium tetani* and *Salmonella enterica* Typhi amongst others (Bacterial vaccines, 2020). However, as with antibiotics, new vaccines are not developed rapidly enough (Pronker *et al*., 2013) due to logistical and immunological challenges. Endemicity of most bacterial infections geographically overlaps with high prevalence of chronic viral (e.g. HIV) and parasitic (e.g. malaria, toxoplasmosis, schistosomiasis) immunomodulatory diseases affecting individuals of all ages (Levine and Farag, 2011), thus rendering the design of safe vaccines against bacterial agents for demographically- and immunologically-heterogeneous populations difficult (Crum-Cianflone and Sullivan, 2017).

Meeting the increasing demand for safe, immunogenic, non-reactogenic vaccines (Khan *et al*., 2017) would be facilitated by a refined understanding of vaccine mode of action at the host-bacterial interface (Coward *et al*., 2014). A variety of vaccine types including live-attenuated, inactivated, subunit and conjugate vaccines is currently available or under development for a wide range of bacterial diseases, each with individual strengths and weaknesses (Clem, 2011). The host's immune response elicited by each vaccine type differs in its immunologic potency, durability, efficacy and cross-protection from relative strains. Additionally, differentially sophisticated technology is required for mass-scale production of each vaccine type, and practical factors including ease of production, chemical resilience under extreme temperatures, route of delivery, cost and frequency of required boosters play a critical role in assessing the eligibility of any given vaccine for mass immunisation in low-resource settings (Duijzer, van Jaarsveld and Dekker, 2018).

It would be ideal if we could produce the cheapest, most practical-to-store and easiest-to-manufacture vaccine that would also confer lifelong immunity against all strains responsible for any particular bacterial disease. However, this is not the case for most available preparations. Different vaccine types have complementary strengths in terms of both their immunological and economic profiles (Clem, 2011), and, ultimately, choices about vaccine type and distribution are made upon factoring in all determinants.

Understanding how, when and for how long each vaccine type impacts on the processes of bacterial replication, killing and inter-organ migration could help identify the minimal functional requirements of an effective vaccine, and potentially inform a synthetic biology approach to vaccine engineering, away from the inefficiencies of empirical design (Diard *et al.*, 2019; Coward *et al.*, 2014),. However, for the majority of vaccines currently in use the mode of action at the level of host-pathogen interactions remains obscure (Oyston and Robinson, 2012).

An understanding of the unobserved processes responsible for the establishment and growth of bacterial populations in different tissues can also render the process of vaccine research and development less wasteful and more systematic. Under the status quo, a vaccine candidate incapable of eliciting the desired immunological response is rejected instead of improved. Largely, this is because knowledge about how a vaccine exerts its *in vivo* effect at the host-bacterial interface is lacking; this knowledge gap hinders the scientific community from making informed modifications to its chemical structure, route of delivery or conjugation before iteratively re-assessing its efficacy (Six *et al.*, 2011).

A good example of how this trend could be reversed is the recent interest in dissecting the effects of imperfect vaccine candidates on the within-host dynamics of pathogens. The merits of this alternative rational vaccine design approach can be demonstrated by the case study of an inactivated oral vaccine against *S.* Typhimurium currently under investigation in the murine host (Diard *et al.*, 2019). Following detailed understanding of the bacteriostatic action of IgA as one of the main protective components of an oral vaccine against *S.* Typhimurium in the gut lumen (Moor *et al.*, 2017), Diard *et al.* (Diard *et al.*, 2019) showed that IgA-mediated bacterial clumping exerts a selective pressure against *S.* Typhimurium, which rapidly evolve surface structures (O-antigens) that can no longer be recognised by vaccine-induced IgA and can, thus, proliferate in the host circumventing the vaccine protection. Insights into this unexpected antibody-driven evolution of IgA escape variants led to iterative improvements in the formulation of the original vaccine, through addition of a cocktail of O-antigen variants that can induce IgA against all observed O-antigen variants.

Finally, characterisation of bacterial migration, killing and division can lead to maximisation of the synergistic potential of bacteriostatic and cidal pathways in bacterial disease management. For example, studying the effect of different vaccine formulations on the early within-host *Salmonella* dynamics has revealed that a live-attenuated vaccine inactivates a large proportion of the inoculum very rapidly, while a whole-cell killed vaccine does not (Coward *et al*., 2014). This rapid inoculum inactivation observed in mice vaccinated with the live-attenuated vaccine is the result of the synergistic action between the vaccine-induced IgG response and the naturally-occurring production of reactive oxygen species (ROS) in the naïve host during the early phase of systemic *Salmonella* infections (Grant *et al.*, 2008). Vaccines and antimicrobial agents that reinforce or complement the immunological processes already in action in the naïve host are likely to be more effective (Lipsitch and Siber, 2016).

1.2. Deciphering the within-host dynamics of bacterial infections

1.2.1. Insights into the within-host dynamics of bacterial disease as a propeller of innovation in management strategies

Previously, it was noted that on the grounds of slow vaccine and antibiotics development, optimal use and repurposing of existing agents is pivotal. A major prerequisite for this strategy is the pursuit of a mechanistic understanding of the biological processes at the host-pathogen interface during infection and treatment.

For the purposes of the current thesis I employ a functional definition of the term *within-host bacterial dynamics* as the collection of the unobserved processes of bacterial colonization, phenotypic switch, division, death and inter-tissue migration within a host following inoculation. Knowledge about the comparative dominance of each process at different stages in the infection timeline and in different tissues can facilitate the transition from empirical to rational-based vaccine and antimicrobial design.

1.2.2. Experimental techniques to investigate interactions at the host-bacterial interface

[reproduced verbatim from Vlazaki, Huber and Restif, 2019]

From first principles, therapeutic interventions should aim to reduce or eliminate the pathogenic bacteria from the infected host. This outcome can be achieved by slowing down bacterial replication, accelerating bacterial killing, altering bacterial migration between tissues or employing any favourable combination thereof. Although bacterial growth and dissemination have been extensively studied in terms of molecular and cellular mechanisms (reviewed by Endesfelder, 2019; Ribet and Cossat, 2015), the efforts to quantify how rapidly these processes progress and change in response to therapeutic interventions remain limited (Levin and Antia, 2001). Quantification of these dynamics requires a detailed observation of within-host bacterial behaviour by means of suitable experimental studies and high-resolution imaging and tracking technologies, as detailed in the following sections.

1.2.2.1. Marker-based methods

The first marker-based methods used to measure the population-averaged rate of bacterial replication *in vivo* were additive, in the sense that they involved an accessory genetic element, including superinfecting bacteriophages (Hormaeche *et al*., 1980; Maw and Meynell, 1968; Meynell, 1959), temperature-sensitive plasmids (Gulig and Doyle, 1993) or plasmids carrying antimicrobial resistance genes (Moxon and Murphy, 1978). In these techniques, genetic elements that induce phenotypic changes are introduced into bacteria and upon division segregate to daughter cells, leading to a reduction in the concentration of the marker, as bacteria divide. The growth of the bacterial population can be modelled with the generation time of the population, signifying the half-life of the marker. At each bacterial generation, we would expect to find one-half fewer bacteria in the population harbouring the non-genetic marker.

Additive marker-based methods have fallen out of favour for higher-resolution experimental techniques, as their output, which is based on detection of phenotypic differences, limits the potential for mathematical inference. Because the concentration of the marker decays with each subsequent bacterial division, the technique is limited to a small number of generations and is only appropriate for studying the early stages of

infection. Additionally, these methods are limited in their ability to study the within-host dynamics of bacteria, as they may be unable to measure inter-tissue migration. Finally, the small number of possible detectable phenotypes constitutes a limiting factor for the resolution of this technique.

To overcome these limitations, non-phenotypic marker-based techniques based on genetic modification of the bacterial core genome were developed, offering the added advantage of unrestricted observational potential in time, as well as tracking of bacterial migration across different tissues. This new generation of marker-based techniques provides outputs that allow the quantification of replication, killing and migration of bacteria *in vivo* (Coward *et al*., 2014; Melton-Witt *et al.*, 2011; Grant *et al.*, 2008; Barnes *et al.*, 2006). In contrast to their predecessors, they introduce modifications to the core bacterial genome, in the form of uniquely identifiable nucleotide tags inserted in non-coding regions of bacterial chromosomal DNA to generate a pool of iso-phenotypic, genetically distinguishable bacteria. This bacterial pool is then inoculated into experimental models. Animals are sacrificed at timepoints of interest, organs harvested, and bacterial composition determined by quantitative polymerase chain reaction (qPCR) or sequencing. Signature-tagged strains allow one to follow the course of infection in the infected host. Because the bacterial are uniquely marked, one can essentially take snapshots of the bacterial loads in different tissues and, using mathematical modelling, determine the rates at which bacteria replicate, migrate and die in tissues at any time during the course of infection.

DNA barcoding experiments have served as an archetype of successfully pairing mathematical with experimental approaches. On many occasions, genetic barcodes have been used in the absence of mathematical models to provide a qualitative assessment of the within-host dynamics of bacterial infections. In a standard infection, observation of a change from a dissimilar to a highly similar composition of tagged populations between two tissue samples would indicate bacterial migration between them. Additionally, the sudden loss of tagged strain diversity would suggest a rapid bottleneck of some effect. For instance, Walters *et al*. (Walters *et al*., 2012) used isogenic tagging experiments to show a bottleneck in the uropathogenic *E. coli* infection between the kidney and the bloodstream.

However, mathematical models further allow for a quantitative assessment of the dynamics of the bacterial infection, as well as model selection. For instance, the observed composition of isogenically-tagged populations of bacteria is the result of the convolved effects of replication, killing, and migration. By using a within-host model which explicitly represents the processes of replication, killing, and migration, it is possible to precisely estimate these rates by comparing model outputs to experimental data. Apart from quantifying the contribution of different processes on the overall bacterial numbers, it is also possible to assess the plausibility of different biological hypothesis by selecting the model with the best fit to the data. Two published studies demonstrate the biological insights offered by complementing the qualitative data interpretation with the quantitative model output. First, Grant *et al.* (Grant *et al.*, 2008) demonstrated using the estimates provided by modelling that in the early stages of infection, replication and killing lead to unique subpopulations of bacteria in different infection foci. Estimation of replication and killing rates would not have been possible without using a mathematical model.

Secondly, Coward *et al.* (Coward *et al*., 2014) introduced elements of mathematical modelling at various points along the experimental timeline. In the early stages of experimental design, simulation-based experiments were used to determine the inoculum dose that would yield the most informative experimental output. Then, mathematical modelling was used to correct the raw experimental data by accounting for the noise introduced due to partial sampling and the samples undergoing qPCR. Finally, mathematical modelling was used to maximise the biological insight by enabling the comparison between the mechanisms of action of the live and killed vaccine. The live vaccine controlled the infection more effectively by inactivating a large number of bacteria shortly after inoculation, and preventing migration of bacteria from the tissues to the blood in the first 48h of infection. In the liver and spleen, mice immunised with the killed vaccine killed bacteria more rapidly than those immunised with the live vaccine. The latter controlled the infection mainly by lowering the replication rates of bacteria in the same organs. Without models, net growth can be estimated from CFU counts, but distinguishing between bactericidal and bacteriostatic effects would have been unfeasible.

1.2.2.2. Fluorescence Dilution-Based Techniques

Fluorescence dilution techniques take the observation of the within-host bacterial dynamics from the population- to the single-cell level (Myhrvold *et al.*, 2015; Claudi *et al.*, 2014; Helaine *et al.*, 2014, Helaine *et al.*, 2010). Fluorescence reporter plasmids, in which dilution of a preformed pool of a fluorophore (fluorescence dilution) acts as a measure of bacterial replication, are introduced into a bacterial population. Flow cytometry is used to determine the intensity of the signal and thus the number of replication events that bacteria have undergone.

By virtue of their single-cell resolution, fluorescence dilution techniques are particularly useful to study bacterial heterogeneity. To this day, mathematical modelling has typically not been used for fluorescence-dilution type experiments. Analyses have focused on qualitatively characterising the heterogeneity in a signal across a bacterial population, e.g. with regards to its spectrum of replication rates.

However, mathematical modelling is needed to precisely quantify the extent of heterogeneity in the biological attribute of interest and to test hypotheses that may underlie this heterogeneity. In theory, because fluorescence halves at each division in fluorescence dilution experiments, discrete peaks should be observed on the histogram of fluorescence corresponding to bacteria at specific numbers of generations. Nonetheless, the histogram of fluorescence intensity is continuous. It is likely that this incongruence reflects some heterogeneity in the experimental process. Mathematical modelling can help distinguish the heterogeneity from the experimental process from the heterogeneity in the biological process. In particular, stochastic models are well-suited for this purpose, because they can capture different sources of variation.

1.2.2.3. Peak-To-Trough Ratio

One of the most recently developed technique pertinent to facilitating the observation of *in vivo* bacterial dynamics makes use of the differential genetic signatures left by replication cycles (Korem *et al.*, 2015). During DNA replication, regions that have already been passed by the replication fork will have two copies, while the yet unreplicated regions will have a single copy. The ratio between DNA copy number near

the replication origin and that near the terminus, termed peak-to-trough ratio (PTR) should reflect the growth rate of the bacterial population.

In 2018, Gao and Li (Gao and Li, 2018) developed an algorithm allowing the quantitative inference of bacterial growth rates from PTR data. First, their approach allows for correction of sequencing bias (GC content bias is reported in next-generation sequencing) using a linear mixed effect models which corrects the average coverage of contig according to the difference in GC content of the average GC content of all contigs. Quality control and exclusion of contigs is followed by a principal component analysis of contig coverages in multiple samples leading to more accurate inference of the distance between the contig and the replication origin compared to single samples. Quantification of parameter estimate accuracy when using experimental data of variable quality (e.g. different number of contig sizes, degree of contamination, contig number) is possible. This feature could be relevant to experimental design, as it could inform experimental biologists of the trade-off between data quality and level of inference accuracy.

1.3. The role of mathematical modelling in maximising insight into the within-host dynamics of bacterial infections

[reproduced verbatim from Vlazaki, Huber and Restif, 2019]

1.3.1. Mathematical Models: Introduction and Classification

Critical as it is to refine our understanding of bacterial disease progression down to the finer scale of host-bacterial interactions, it constitutes a challenge despite the increasing availability of new experimental techniques. With prospectively informed experimental design and improving bacterial tracking techniques in the lab, dynamical mathematical models have the power to infer the parameters governing the hitherto unobservable processes underlying bacterial dynamics, quantify them and characterize the uncertainty around them.

In 1974, Eykhoff (Eykhoff, 1974) defined a mathematical model as "a representation of the essential aspects of a system, presenting knowledge of that system in a usable form". These mathematical "representations" of real-life systems are equations organised under a range of forms including boolean models, models based on ordinary or partial differential equations, stochastic models based on stochastic differential equations, agent-based models and game theory models amongst others (Cantone *et al.*, 2017; Ciupe & Heffernan, 2017) in what Wolkenhauer wittily referred to as "a zoo of modelling approaches" (Wolkenhauer, 2014). What makes all animals in that zoo distinctly powerful is their ability to address very different questions one may have about a system

Though microbiologists are familiar with experimental models, they are traditionally less familiar with the versatility of mathematical modelling approaches, largely due to the limited interdisciplinary training in biological sciences (Levin and Antia, 2001). In this section, we aim to provide a convenient introduction to mathematical models by classifying modelling techniques according to their comparative strengths and applications to microbiological questions.

Mathematical models can be categorised into functional classes (Figure 1.1), according to different criteria such as (1) whether their parameters represent biological processes, (2) whether they are fitted to experimental data, (3) whether molecular, cellular or other sources of variability affect the output of the model, (4) what the purpose of model development is.

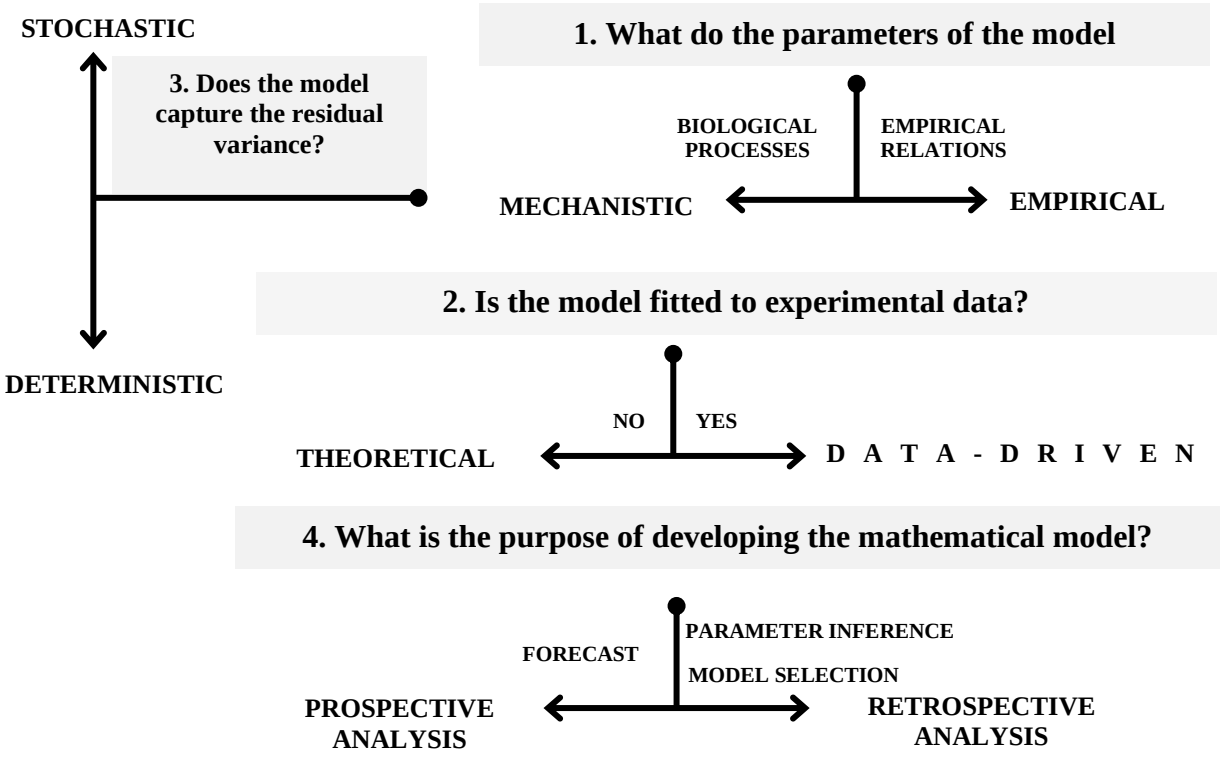


Figure 1.1: Schematic representation of the relationships between different mathematical modelling techniques *[reproduced from Vlazaki, Huber and Restif, 2019]*

Mathematical models can be divided into mechanistic and empirical on account of whether their parameters represent biological processes or simply characterise relations between variables in the data. While empirical models are necessarily data-dependent, mechanistic models can either be system-specific and fitted to data or, generic, explorative and not related to experimental data (theoretical). Additionally, depending on the biological question addressed, mechanistic models can be deterministic when only the average behaviour of the system is of interest, or stochastic when unexplained variation in the behaviour of the system matters too. Mechanistic data-driven models can serve different purposes: they can either be solved forward in time to make a forecast (prospective analysis) or can be solved backwards in time to perform parameter inference and model selection (retrospective analysis).

1.3.1.1. Stochastic, mechanistic data-driven models

In this thesis I use stochastic mechanistic models coupled with experimentally observed data. Mechanistic models incorporate the biological mechanisms by which changes in the system are thought to occur and require some knowledge or speculation about the unobserved interactions that determine the observable output (Thakur 1991; Baker *et al.*, 2018). In the context of mathematical modelling, a mechanism often does not correspond to the molecular, cellular or genetic hierarchy that collectively constitutes causality from a microbiologist's viewpoint. From a modeller's point of view, a mechanism is a conceptual representation of the process acting to change the state of the system in a certain direction. Structurally, mechanistic models can be represented with flow diagrams, whereby the state of the biological system, i.e. an imaginary complete snapshot of that system at a given point in time, changes through processes measured by parameters in the model (Figure 1.2). Examples of biologically relevant quantities encoded as variables in these models (i.e. boxes in the flow diagram) can include concentrations of cytokines and antibodies, as well as numbers of immune cells and infectious agents in tissues amongst others. These quantities are at times directly measurable experimentally and at others only determinable by proxy.

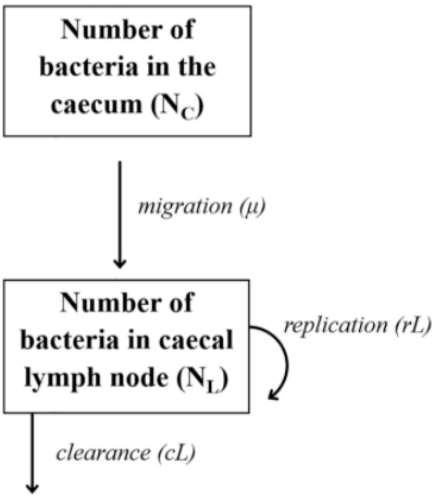


Figure 1.2: An example of a flow diagram as a schematic representation of mechanistic models in microbiology *[adapted from Kaiser et al., 2013]*

Boxes represent the variables of the system, in this case the number of bacteria in the caecum and caecal lymph node respectively. Mathematically these variables are shortened for convenience as N_C and N_L. The processes that change the state of the system are represented by arrows and, here, correspond to bacterial migration, replication and clearance. The rates at which these processes take place are quantified by parameters, in this case, μ, rL and cL corresponding to the rates of migration, replication and clearance, respectively. *[reproduced verbatim from Vlazaki, Huber and Restif, 2019]*

Mechanistic data-driven models, in particular, are usually system-specific and parameterized based on experimental outputs. For example, in 2014, Coward *et al.* carried out an experiment using individually tagged strains of *S.* Typhimurium to determine the effects of different vaccines on the rates of replication and killing of bacteria. The measurements of bacterial numbers in the differentially tagged subpopulations along the infection timeline were fed into a population-based mathematical model, which permitted estimation of the rates of replication and killing of bacteria under the two immunization regimens enabling the direct comparison between them.

Stochastic models yield different results each time when initialized with the same parameters and initial conditions. While multiple realizations of stochastic models recapitulate the range of potential outcomes and their likelihood of occurrence, the output of a deterministic model corresponds to the mean outcome of these realizations. Stochastic models can quickly become too complex to solve analytically, and their exploration may, thus, be entirely simulation-based. When stochastic processes are simulated, the random nature of trajectories generated is due to randomly, exponentially distributed, picked times until the next process, with a randomly picked process executed during that time (Gillespie, 1977). This can become computationally expensive.

Despite the higher computational cost of stochastic models, they often constitute the only reasonable modelling choice, when the behaviour of the system in question is particularly influenced by stochasticity. With terminology borrowed from the field of ecology, the uncertainty in the outcome can be decomposed as a function of two sources of stochasticity: demographic (e.g. Shaffer, 1981; Burgman, Ferson and Akcakaya, 1993) and environmental. Demographic events include births, deaths and migration of individuals. The mean rate at which a demographic event occurs is defined as the inverse of the average time it takes for the event to take place and can be quantified with mathematical models. These events are described by binary random variables with a certain probability of occurring per given unit of time. As demographic events at the population-level are a function of the sum of demographic events at the individual level, the strength of demographic stochasticity is greater for small populations (Kokko and Ebenhard, 1996).

On the other hand, environmental stochasticity is independent of the individual; rather, it refers to unpredictable changes in the environmental conditions that the individuals experience. As such, its effects do not depend on the population size, but on the number, heterogeneity and stability of factors influencing individual behaviour (Fujiwara and Takada, 2017). In the context of within-host infectious dynamics, the nature of the pathogen determines the strength of each source of stochasticity. For infections requiring a small founding population, demographic stochasticity becomes important as they may either be successful or become extinct. At the same time, the complexity of the immune response, the microenvironment of the tissue(s) colonized, and the heterogeneity of infectious foci determines the effect of environmental stochasticity on the population. When prior biological knowledge indicates that either source of stochasticity may be strong, it is advisable to use stochastic rather than deterministic models.

1.3.2. Examples showcasing the range of biological insights gained using mathematical models in the field of host-pathogen interactions

[reproduced verbatim from Vlazaki, Huber and Restif, 2019]

1.3.2.1. Quantification of the relative contribution of different immune system components to the progression of infection

Within-host models of parasitic, viral and bacterial infection often seek to determine the relative contributions of different components of the immune system in regulating the dynamics of infection. In the current treatment paradigm, the role of the host's immune response is often neglected, and therapeutic agents are administered for fixed periods of time usually in the form of monotherapy and regardless of the infectious load. It is now becoming increasingly recognized that a first step toward optimization of existing therapies is the induction of synergistic effects between the host immunity and the standalone effect of the therapy (Gjini and Brito, 2016). To thoroughly understand this interaction, mechanistic mathematical models can be used in two main ways. First, one can use a series of nested models prospectively, starting from simpler models and adding features of the immune response while quantifying the impact of each new addition in the process. Second, by enabling the segregation between unobserved processes of replication and killing, mathematical models begin to shed light on the black-box of host–pathogen interactions and inform further biological experiments. For example, if rapid killing is identified as the main driver of an observed decline in bacterial numbers, it is reasonable to first look in the direction of known cidal branches of the host immunity.

One such approach was taken by Grant *et al.* in 2008 (Grant *et al.*, 2008) where following the identification of early bactericidal activity in a *Salmonella* mouse model, infection progression in wild type mice was compared to that in NADPH oxidase deficient mice to unravel an important role of that immunological component in inducing the inferred bactericidal effect. In other pathogen classes, mathematical models addressing similar questions were successfully used much earlier.

With regards to host immune system-bacterial interactions, mathematical modelling of the *Mycoplasma* species has been ongoing. The Kirschner group have developed a series of increasingly complex mathematical models to describe the role of different cell types

and chemokines of the immune system in the progression of early tuberculosis (TB) infection. Their compartmental model based on ordinary differential equations including the lung and the draining lymph node (DNL) has been used to study the dynamics of early infection, particularly the role of dendritic cells in T-cell priming (Marino and Kirschner, 2004) and, later, the roles of dendritic cell trafficking to and from the DNL (Marino *et al.*, 2004), cytotoxic T-cell-mediated *Mycobacteria* killing (Sud *et al.*, 2006), TNF-α and anti-inflammatory IL-10 (Cilfone *et al.*, 2013) in host defence. Their contributions identified the macrophage infection rate and T-cell-mediated immunity as the two key elements in determining the trajectory of an infection into one of (i) primary TB, (ii) primary TB with clearance, (iii) latency and (iv) reactivation (Marino and Kirschner, 2004).

1.3.2.2. Comparison of the effect of different strains on infection dynamics

Fitting within-host models to samples of different strains of the infectious pathogen can also facilitate our understanding of how the within-host dynamics of infection vary across different strains of the same species. Different strains of pathogens are responsible for differences in seasonal and local outbreaks of contagious and deadly infections such as influenza (Du *et al.*, 2017), cholera (Weill *et al.*, 2019), community-acquired pneumonia (Zhang *et al.*, 2019) and others. These pathogens, albeit very closely related, can show extreme differences in transmission rates (e.g. in *Mycobacterium tuberculosis* in Verma *et al*., 2019), response to therapeutic agents (e.g. in *Vibrio cholerae* in Weill *et al.*, 2019) and virulence (e.g. in swine fever virus in Portugal *et al.*, 2014). In this context, mathematical models allow for sensitivity analyses to identify which parameter(s) have the greatest impact on a given outcome; this can be helpful in highlighting potential causes that drive inter-strain differences.

For instance, Hur *et al.* (Hur *et al.*, 2013) fit models of influenza infection to experimental data on seasonal and pandemic strains of flu. They found that the only parameter that varied between the pandemic and seasonal strains was the viral replication rate, indicating that intracellular viral replication may affect pathogenicity.

1.3.2.3. Comparison of the effects of different therapeutic interventions on infection dynamics

Within-host models of infection (both theoretical data-driven) can also reveal important insights into the effect of different drugs at the level of the host–pathogen interaction and identify effective treatment strategies (e.g. decide whether it is more efficient to prevent replication or increase killing). For example, Rong and Perelson (Rong and Perelson, 2010) evaluated the effect of different Hepatitis C (HCV) treatment strategies. Protease inhibitors are being increasingly used in combination with pegylated interferon and ribavirin to treat HCV-1 infection, but there remain concerns of relapse after treatment. They developed a deterministic mathematical model to examine viral load dynamics before and after treatment with a protease inhibitor. Banerjee, Keval and Gakkhar (Banerjee, Keval and Gakkhar, 2013) considered the effect of ribavirin being used in combination with interferon therapy for HCV infection. Although the study was theoretical in nature, it found that – provided a certain threshold of drug efficacy – a triphasic response of viral load could be observed, leading to eradication of the virus.

1.3.2.4. Comparison of the effects of different inoculum size on infection dynamics

Infections can take off with inocula of variable sizes. However, the inoculum size affects the population composition of the infectious agents and how they respond to therapy. Formulating a deterministic mechanistic model, Meredith *et al.* (Meredith *et al.*, 2015) reported that inoculum size determines the efficacy of β-lactam antibiotics when administered to bacterial populations of which at least some members harbor extended β-lactamase activity. If β-lactam antibiotics were administered in high-density populations, then some members would survive and re-establish the infection. They reported that the population was sensitive when its initial density was sufficiently low or examined in a short time window. Given these properties, they reasoned that optimal antibiotic dosing may remain effective in bacterial populations even when they harbour resistance genes.

1.3.2.5. Studying the dynamics of infection across different scales

Mathematical models can be employed to study host–pathogen interactions at multiple levels, from cellular to whole-organism and even population level. A solid knowledge of the versatility of mathematical techniques allows the use of the same tools to study questions on different scales. With judicious use, mathematical models can also combine insight acquired at different levels e.g. the single cell and organ levels and use this to gain novel insights about disease progression (Gog *et al.*, 2015). For example, a stochastic mathematical model generated by the Perelson lab showed that early HIV dynamics differ depending on whether infected target cells produce virions continuously or do so in a single burst (Pearson, Krapivsky and Perelson, 2011). This study shows how events at the single-cell level can have a profound impact on infection dynamics at the whole-organism level ultimately affecting clinically important quantities used for diagnosis and as guides for therapeutic intervention.

Furthermore, it is possible to use the predictions from modeling the host–pathogen interactions to inform models at higher scales. In 2009, Heffernan and Keeling (Heffernan and Keeling, 2009) took advantage of well-founded predictions about immunity in a measles-infected host (Heffernan and Keeling, 2008) to predict the effect of vaccination at the population level.

1.3.2.6. Investigating the evolutionary dynamics of infectious disease within the host

Finally, mathematical models have been used to characterize and quantify the evolutionary dynamics of infectious agents within a host. For instance, Chisholm and Tanaka (Chisholm and Tanaka, 2016) developed a mechanistic mathematical model to examine the evolution of *M.* tuberculosis within its host. *M.* tuberculosis is observed to enter a latent, dormant state, but, at first glance, a state of dormancy is not advantageous for the pathogen as it does not permit replication. However, the study demonstrated that latency can be an evolutionarily desirable state.

Furthermore, Fabre *et al*. (Fabre *et al*., 2012) formulated a deterministic mechanistic model of competing viral populations within host plants. They parameterized it according to the carrying capacity of the plant, the intrinsic rate of increase of each variant and the competition strength each genotype exerts on the others. They determined the forms of selection processes occurring between competing viruses within a host plant, and the

intensity and temporal variation of genetic drift experienced by viruses during host plant colonization. Parameters were statistically inferred by model fitting to high-throughput sequence data of the viral counts obtained from the plants over time, and model selection was performed (after testing several models reflecting different mechanisms of competition).

1.3.3. Inference frameworks

As noted earlier, mathematical models can serve a range of purposes (Figure 1.1) from forecasting to inferring parameters and aiding in experimental design. In data-based mechanistic models the first step to achieving any of these goals is model calibration, a process of identifying combinations of parameter values that best align the model's output with the experimentally observed biological output (Read *et al.*, 2018). Despite its central role in data-based mathematical modelling, model calibration is often challenging for biological systems with multiple components, complex underlying biological mechanisms and long progression timelines due to various limitations in inference frameworks (Penas *et al.*, 2017). These issues are particularly relevant to *in vivo* studies on the within-host dynamics of infection where the capacity for observations and direct measurements is limited.

An inference framework is a validated set of mathematical tools aiming to identify a single set or a range of parameter values that either minimise a defined metric of discrepancy or maximise a metric of proximity between an experimentally observed and a model predicted output. In this section, I explain the motivation for my inference framework of choice by classifying commonly used inference frameworks into a) likelihood-based (statistical) and b) likelihood-free (simulation-based), and outlining the features, advantages and limitations of each.

1.3.3.1. Likelihood-based inference frameworks

Likelihood-based frameworks notably include frequentist maximum-likelihood (MLE) and Bayesian frameworks (Hartig *et al.*, 2011). Both rely on the idea that $P(x|\theta)$, the probability of obtaining the observed data x for each possible model parameterisation θ given model $M(\theta)$ is proportional to the support given to a parameter set θ by the observed data x (Fisher, 1922).

On this premise, the frequentist MLE approach formulates a likelihood function $P(x|\theta)$ and seeks to maximise that function; the θ that corresponds to the maximum likelihood value is taken to be the best point parameter estimate for a given dataset x.

In contrast, the Bayesian framework is interested in a compound quantity known as the posterior distribution $P(\theta)$ which is proportional to the likelihood $P(x|\theta)$ multiplied by the prior distribution $p(\theta)$, a probability distribution reflecting one's prior belief about the parameter values prior to model calibration (Equation 1).

$$P(\theta) = p(\theta|x) = \frac{p(x|\theta)\cdot p(\theta)}{\int p(x|\theta)\cdot p(\theta) d\theta} \qquad (1)$$

Unlike the frequentist approach which identifies a single best point estimate for the underlying parameters, $P(\theta)$ (Eq. 1) is interpreted as a distribution of parameter values reflecting the uncertainty around the parameter values in the Bayesian framework.
Strictly speaking both frameworks are contingent upon the formulation and solution of a likelihood function $P(x|\theta)$. While for some simple models it is indeed possible to solve the likelihood function analytically, most models have *intractable* likelihood expressions which require numerical approaches.

1.3.3.2. Likelihood-free inference frameworks

Although analytical solutions are the fastest and most accurate, for the vast majority of biologically relevant mathematical models the likelihood expression is impossible to evaluate due to the presence of a large number of interacting components, multiple latent variables, or a distribution of observed data that is modelled by a complex stochastic computational model.

To bypass this challenge, it is possible to use a *forward* approach starting with parameter values, followed by model-based simulation and comparison of the simulation output x' to the observed data x. This simulation-based, likelihood-free approach constitutes the basis for a range of inference frameworks (reviewed in Durham and Gallant, 2002) including approximate Bayesian computation (aBC), (Marin *et al*., 2011; Beaumont, 2010; Csillery *et al.*, 2010), simulated pseudo-likelihoods (Concordet and Nunez, 2002), indirect inference (Gourieroux *et al*., 1993), informal likelihoods (Beven, 2006), sequential neural likelihood (Papamakarios, Sterratt and Murray, 2019), divergence-minimisation (Price *et al*., 2017) and Pattern-Oriented Modelling (Wiegand *et al.*, 2003, Grimm *et al*., 2005). These tools rely on approximating the shape of the likelihood function by means of efficient sampling and optimisation techniques often combined with

reducing the dimensionality of the data by comparing sufficient summary statistics between the observed and simulated data (Hartig *et al*., 2011).

1.3.3.2.1. Divergence-minimisation inference framework

In 1.3.3.2. I list only a selection of the numerous likelihood-free methods available to overcome the challenge of intractable likelihood expressions. The impetus for the development of this large number of inference tools has been the quest for computational efficiency as simulation-based tools have high demands in computational time and power. However, despite the range of available tools, three practical problems persist:

1. computational efficiency remains suboptimal, often only allowing for parameter inference given a single model, but rarely allowing for model selection where multiple models need to be parameterised,
2. the majority of tools available have been developed and applied in disciplines such as physics, engineering and computer science; as a result, rarely are there blueprints of applied case studies accessible to biologists,
3. in microbiology, the format of the observed data output is inflexibly dependent on the resolution of the experimental techniques used, and any inference framework used needs to be adjusted and optimised for the format of data in question.

As a result of these unresolved challenges, the range of suitable inference tools is, in reality, limited. My methodology of choice in this thesis has been a moments-based, divergence-minimisation approach developed by Price *et al.* (Price *et al.*, 2017), which has been applied to ITS-based data before and has shown to be computationally efficient for model comparison. Its computational efficiency results from a combination of a) dimensionality reduction by compressing the experimentally observed and model outputs into summary statistics (method of moments), and b) obtaining analytically evaluable expressions of the first- and second-order moments used as summary statistics for the linear dynamics assumed in the model.

In this framework, the metric used to quantify the degree of support that observed data provide to a given model M given a set of parameters θ is the Kullback-Leibler (KL)

divergence between the summary statistics of the model output and those of the observed data. Instead of seeking to maximise the likelihood of observing the experimental data under a parameter set θ, here one seeks to minimise the discrepancy between the distribution of data generated by $M(\theta)$ and that of experimentally observed data. Confidence intervals around the parameters point estimate are obtained by means of parametric bootstrap followed by parameter estimation for each of the bootstrapped samples, as detailed in section 2.7.

Overall, the computational efficiency of this moments-based likelihood-free inference framework facilitates the application of mathematical models to *in vivo* observations from complex biological systems over a long time period, the comparison of the performance of multiple candidate models, model selection and the use mathematical modelling as a tool for optimising experimental design.

1.4. Murine infections with *Salmonella enterica* serovar Typhimurium as an *in vivo* experimental model

1.4.1. Background on the experimental system

Experimental models are physical representations of a real-life system either *in vivo* or *in vitro*, commonly used by microbiologists in the lab because of their amenability to genetic manipulation, reproducibility of results and convenience. While *in vitro* work can reveal mechanistic insights at the molecular and genetic level in terms of pathogenicity and response to treatment, *in vivo* work with animal hosts introduces additional layers of biological complexity and renders findings more relevant to whole-organism biology. All animal models of human diseases have inherent limitations, but they also have important advantages over *in vitro* methods, including the presence of organized organ systems, an intact immune system, the possibility to engineer animal hosts with specific genetic backgrounds, and the availability of multiple methods for characterising the immune response to microorganisms (Bakaletz, 2004).

Murine infections with *Salmonella* species have been widely used as an experimental model to study the bacterial pathogenesis of gastrointestinal diseases and host-pathogen interactions *in vivo* (Higginson, Simon and Tennant, 2016). This host-pathogen combination is popular because, on the one hand, *Salmonella* is an extensively characterised pathogen with relevance to human diseases and, on the other, because certain mouse strains can be readily infected by *Salmonella* and show symptoms within a relatively short timeframe. The abundance of knowledge around the pathogen and the susceptibility of the host to infection, alongside other logistical facilities such as standardised ethics protocols, low-cost housing requirements and easily accessible rodent handling expertise make murine *Salmonella* infections an attractive choice as an experimental model in vaccine and therapy design, as well as immunological and microbiological studies (Higginson, Simon and Tennant, 2016; Nilson, Kari and Steele-Mortimer, 2019).

1.4.2. Systemic Salmonellosis: Epidemiology, Challenges and Management

Besides its central role as a prototypical experimental system, murine infections with *S.* Typhimurium have been extensively studied as a model system for both the enteric and systemic phase of salmonellosis. Systemic salmonellosis refers to a collection of human systemic diseases caused by different strains of the same bacterial pathogen; it encompasses typhoid and paratyphoid fever caused by infection with *Salmonella enterica* serovars Typhi and Paratyphi respectively, and **i**nvasive **N**on-**T**yphoidal **S**almonellosis (iNTS) caused when non-typhoidal intestinal *Salmonella* strains (Enteritidis, Typhimurium and thousands more), which can escape the host's immune system control and disseminate into the blood (Feasey *et al*., 2012).

Salmonella enterica is a Gram-negative enteric bacillus, commonly transmitted through the faeco-oral route and less frequently directly into the blood through an open wound. Upon ingestion, the bacterium successfully survives the low gastric pH and moves to the intestine where it invades epithelial cells ultimately gaining access to the underlying lymphoid tissue. Abundant with macrophages, the lymph facilitates the encounter of phagocytes with *Salmonellae*, which henceforth establish an intracellular niche used for proliferation, immune evasion and spread. Systemic infection arises at the lymph – blood interface where the establishment of bacteraemia facilitates bacterial migration and colonisation beyond the mesenteric and caecal lymph nodes to the systemic compartment with the liver and spleen being the main reservoirs in the murine hots (Parry *et al*., 2002; Grant *et al*., 2008)

Although largely treatable with antibiotics and preventable by improvements in water hygiene and general sanitation, systemic salmonellosis was still responsible for 14 million cases of typhoid fever with 222,000 associated deaths, over 2 million cases of iNTS with >500,000 associated deaths (Ao *et al*., 2015), and unaccountable cases of neuropsychiatric sequelae in survivors in 2014-15. The geographical distribution of disease burden is uneven, with South-East Asia and Sub-Saharan Africa being regions of typhoid fever and iNTS endemicity (Figure 1.3) mainly due to economic inequality, poverty and inadequate sanitation infrastructure, coinfections and high HIV prevalence. To make matters worse, reports suggest that a large proportion of those outbreaks included cases of multidrug-resistant infections unresponsive to mainstream antibiotic monotherapy administered under traditional timeframes (Yan *et al*., 2016; Gordon *et al*.,

2008), rendering antimicrobial resistance an additional, clinical obstacle in the reduction of worldwide mortality and morbidity of systemic salmonellosis.

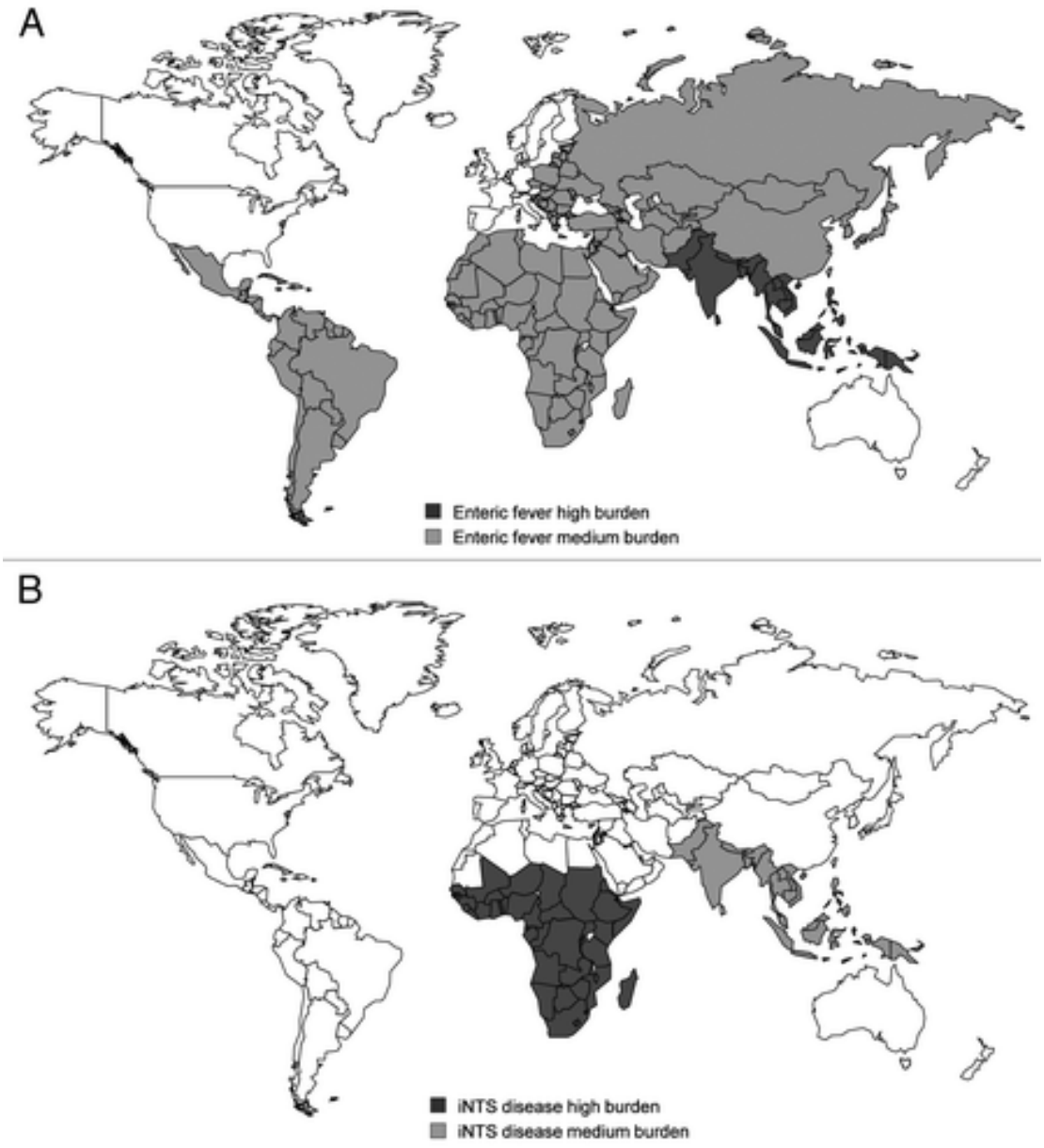


Figure 1.3: Geographic distribution of typhoidal and non-typhoidal *Salmonella* endemicity
[reproduced from MacLennan, Martin and Micoli, 2014]

Chronic asymptomatic human carriage further complicates the landscape of transmission and bacterial evolution; acutely infected patients who clinically recover can remain either short-, longer-term or chronic carriers responsible for human-human disease transmission through faecal shedding. They constitute living and moving disease reservoirs facilitating the emergence of novel bacterial genotypes (Gunn *et al*., 2014) through horizontal transfer of antimicrobial resistance genes. Around 5% of those infected and treated for salmonellosis will proceed to chronic bacterial carriage (Levine, Black and Lenata, 1982) in the biliary tree and gallbladder through hepatobiliary bacterial transfer (Gonzalez-Escobedo, Marshall and Gunn, 2011). Typhoidal *Salmonella* strains have no known environmental reservoirs and shedding from a chronic, asymptomatic carrier may be a key component in driving transmission in human populations (Gunn *et al.*, 2014). Bacterial persistence is also critical to address from the viewpoint of effective antimicrobial stewardship: can extended antimicrobial regimes ever eradicate these infections and if so, in what combination and length?

There are at least two explanations about why antibiotics, even with optimal use, may fail to completely clear a bacterial infection. The first is that bacterial targets are resistant to the antimicrobial in use, i.e. they carry inheritable and transferrable genetic elements allowing them to resume grow at high antibiotic concentrations (Scholar and Pratt, 2000). The second explanation is the phenomenon of tolerance whereby some bacteria survive exposure to antibiotics because of their slow metabolic state while remaining genetically identical to bacteria susceptible to the administered antimicrobial (Brauner *et al.*, 2016). Tolerance naturally ensues from non-genetic, phenotypic heterogeneity, defined as an epigenetically-driven metastable variation in cellular processes (Dhar and McKinney, 2007). In turn, heterogeneity in bacterial populations has multiple putative sources ranging from deterministic in origin and including cell age, cell-cycle stage and metabolic oscillations to stochastic events related to gene expression (Avery, 2006). Focusing on stochastic determinants of heterogeneity, a bidirectional causal relationship between heterogeneity in bacteria and heterogeneity in host cell response further amplifies the range of observed host-pathogen interactions (Avraham *et al*., 2015).

Appreciating the reciprocal relationship between bacterially- and environmentally-driven population heterogeneity can contribute to antibiotic use rationalisation. For example, unveiling the specifics of non-genetic heterogeneity in bacterial replication

rates could favour the use of antibiotics in polytherapy regimes instead of standalone therapy or vice versa depending on their predicted synergistic or antagonistic effects. Furthermore, finding out when this phenotypic heterogeneity is first established *in vivo*, whether bacteria with different replication rates follow a characteristic spatial distribution and whether tissues with specific features preferentially act as reservoirs has the potential to optimise the route of antimicrobial delivery and its formulation, and to inform best clinical practice about whether prolonged antibiotic regimes can ever succeed in fully clearing the bacterial reservoirs.

Currently, the clinical armamentarium against systemic salmonellosis is based on the pillars of antibiotic therapy and pre-exposure immunisation. Public health authorities in the U.K. recommend the use of broad-spectrum antibiotics such as fluoroquinolones (e.g. ciprofloxacin, orofloxacin), chloramphenicol or ampicillin as first-line treatment of sensitive clinical isolates of *Salmonella*, administered for 7-14 days. In cases of long-term carriage, often seen in immunocompromised individuals, 28-day antibiotic courses are recommended, albeit with equivocal success in achieving complete bacterial clearance (British Medical Journal (BMJ) Best Practice, 2018). The emergence of *S.* Typhi strains resistant to all first- and second- line treatment antibiotics (including ampicillin, trimethoprim-sulfamethoxazole, and chloramphenicol) has increased in frequency worldwide since the 1980s, creating pressure to shift the management paradigm towards prevention rather than treatment in endemic areas (Kariuki *et al*., 2015). Alarmingly, in Malawi, 90% of iNTS isolates were multi-drug resistant in 2008 (Gordon *et al.*, 2008).

Although prevention strategies against systemic salmonellosis have been an active area of pharmaceutical research and development since the 1890's, only two widely-available vaccines against *S.* Typhi have been licensed to-date (MacLennan, Martin and Micoli, 2014), and no vaccines are available against other *Salmonella* serovars responsible for paratyphoidal and non-typhoidal salmonellosis (Coward *et al.*, 2014; MacLennan, Martin and Micoli, 2014). Live (e.g. Ty21a), inactivated polysaccharide-only (Vi capsular polysaccharide) and conjugate protein-polysaccharide vaccines against *S.* Typhi are currently in use (Higginson, Simon and Tennant, 2016), the second confirmed as safe for administration to immunocompromised individuals (World Health Organisation (WHO), 2018). However, although licensed, the vaccines remain underused due to their poor

immunogenicity in young children (MacLennan, Martin and Micoli, 2014). The increasing paratyphoidal salmonellosis prevalence in endemic countries (Ochiai *et al.*, 2005; Woods *et al.*, 2006) and the >20% fatality rate of non-typhoidal salmonellosis amongst children (Tapia *et al.*, 2015) renders the need for vaccines conferring protection against a wider range of *Salmonella* serovars pressing.

The efficacy of licensed vaccines to induce long-term immunity has been traditionally explored from the viewpoints of field epidemiological surveys and lab-based molecular immunology, and only recently have mathematical models been used to interrogate the within-host dynamics of infection, in what ways and how early in the infection timeline different vaccines influence bacterial killing, replication and inter-organ migration (Coward *et al.*, 2014). Equally, persistence, chronic asymptomatic carriage and relapse of infection have long been studied from a clinical, microbiological and genetic point of view, but to decipher why these phenomena occur *in vivo* whilst the bacterial reservoir retains its *in vitro* susceptibility to antimicrobials requires an understanding of how the spatial distribution of bacteria affects antibiotic efficacy as infection progresses, how common antimicrobial classes affect the bacterial rates of division, killing and migration, and, finally, when and at what rate heterogeneous pathogen subpopulations arise in the tissues (Rossi *et al.*, 2017).

Insights into the dynamical processes that drive bacterial population fluctuations at the host-bacterial interface has been arising as a potential accelerator of the empirical-to-rational transition in disease management.

1.4.3. Literature review on the within-host dynamics of salmonellosis

In 1.4.1. it was noted that *Salmonella* infections of the murine host is a reliable and widespread used experimental model for the study of host-bacterial interactions *in vivo*. As such, it has formed the basis for a range of qualitative and quantitative studies with the aim to characterise the within-host dynamics of *Salmonella* behaviour in the naïve host, differentiate between the effects of interventions at the host-pathogen interface, study the behaviour of *Salmonellae* under antibiotic pressure in different tissues, and identify the salient immunological components governing each phase of the infection. This section provides a thematic categorisation of the existing literature based on the biological questions addressed by these studies with a special focus on *Salmonella enterica* serovar Typhimurium (Table 1.1). These questions range from exploratory research mapping out the bottlenecks, intra- and inter-tissue dynamics, identification of relevant host immunity control mechanisms to hypothesis-driven research comparing the effects of two or more interventions on the dynamics of infection, and others.

1.4.3.1. Characterisation of colonisation bottlenecks, intra-organ growth and inter-tissue migration

In the field of population dynamics microbiology, bottlenecks refer to events that reduce the size of the bacterial population and typically take place when a pathogen first colonises a new host, upon migration to a different tissue, under immune pressure, and upon treatment with antibiotics (Abel *et al.*, 2015; Rossi *et al.*, 2020). Intra-organ growth is the net bacterial growth that underpins the observable number of bacteria in a tissue and represents the convolved effects of local bacterial replication and death, while inter-tissue migration describes the process of bacterial movement between tissues or organs. Studying these processes can help to a) characterise the timeline and directionality of tissue colonisation during the infection and adjust the timing and target of therapeutic interventions accordingly, b) take advantage of the spatiotemporal dynamics of infection to potentiate the effects of immunisation or administered antibiotic therapy and limit the emergence of resistance, c) understand the complex interplay between different bacterial species in co-infections or at the host microbiome-pathogen interface (Abel *et al.*, 2015; Coward *et al.*, 2014).

Studies using murine infections with *S.* Typhimurium have used a range of experimental techniques to investigate the bacterial dynamics in the naïve host in both the enteric (Kaiser *et al.*, 2013) and systemic phases (Grant *et al.*, 2008) of infection, as well as at the enteric-systemic interface (Lam and Monack, 2014; Gulig and Doyle, 1993). Kaiser *et al.* (Kaiser *et al.*, 2013) used ITS-based data combined with mathematical modelling to show that bacterial migration from the gut to lymphoid tissue is almost entirely unidirectional and constitutes a major bottleneck in the progression of infection. Earlier, Grant *et al.* (2008) had shown that in the systemic phase of infection high cidal activity imposes a strong bottleneck with concomitant rapid bacterial replication leading to the establishment of independent bacterial foci in different organs.

The progression of *S.* enterica from the enteric to the systemic phase is of particular clinical interest, as it is commonly associated with systemic salmonellosis in immunocompromised and elderly individuals. Three ITS-based studies converged to the conclusion that salmonellae are subject to a strong bottleneck effect in their transition from the enteric to the systemic phase, leading to a pronounced founder effect in the colonisation of the mesenteric lymph node, liver and spleen (Kaiser *et al.*, 2013; Kaiser *et al.*, 2014; Maier *et al.*, 2014). Gulig and Doyle (Gulig and Doyle, 1993) narrowed down the virulence activity of the *S.* Typhimurium plasmid to a 6.2-kb genomic region and found that its presence increases the rate of intracellular bacterial replication without affecting bacterial death or migration. Finally, Lam and Monack (Lam and Monack, 2014) studied the effect of orally and intraperitoneally acquired *S.* Typhimurium co-infections and concluded that the enteric *Salmonellae* exert intraspecies priority effects, leading to the exclusion of the intraperitoneally inoculated strain from further colonisation.

1.4.3.2. Identification of host immunity mechanisms responsible for infection control

Once a pathogen colonises a new host, the outcome of the infection is multi-factorially shaped by the interaction between the pathogen's virulence factors and the host's immune responses, compounded by demographic and environmental stochastic factors including variation in the bacterial replication, death and migration events as well as the composition of the humoural and cellular microenvironment at the time (Vlazaki, Huber and Restif, 2019). While this stochastic environment is ever-changing, the majority of host-pathogen interactions are conserved and can be demystified with the aim of

developing improved diagnostics, therapeutics, and prophylactics (Thakur, Mikkelsen and Jungersen, 2019; Maier *et al.*, 2014; Grant *et al.*, 2008).

In the quest for conserved patterns of host-pathogen interactions, a large body of work in microbiology pertains to genetic and bioinformatics studies identifying virulence factors and host immunity mechanisms by predicting the function of genes through their genetic sequence, using transgenic mice and DNA microarrays (Falkow, 2004). In parallel, studies in the field of population dynamics microbiology have successfully identified such host-pathogen interactions by identifying dominant cidal or static patterns, deducing which immunological factors may be responsible for the observed effect, and ultimately verifying hypotheses by comparing the dynamics of the same infection in knockout mice lacking the gene encoding for the immunological component in question. Genetic and population dynamics studies frequently complement each other to pinpoint the gene(s) behind an immunological mechanism of interest.

An elegant example of the intertwined relationship between the two fields is found in the experimental system of murine *S.* Typhimurium infections. Early genetic studies had shown that while some mouse strains are resistant to *S.* Tyhpimurium, others are susceptible (Plant and Glynn, 1976); the mechanism remained unknown. In a follow-up population dynamics study, Hormaeche (Hormaeche, 1980) noted that *Salmonellae* grew at almost double the rate in susceptible BALB/c compared to resistant (B10 x A/J)F1 mice, and proceeded to show that the difference in net growth was entirely attributable to higher bacterial division rate in the former from the early stages of infection, possibly in macrophages. In a subsequent genetics study, Vidal *et al.* (Vidal *et al.*, 1995) pinpointed the natural resistance-associated macrophage protein *Nramp1* gene on the *bcg* locus as the causative agent for the difference in susceptibility between the mouse strains.

Furthermore, Grant *et al.* (Grant *et al*., 2008) used an ITS-based approach in phox$^{-/-}$ mice, combined with mathematical modelling to show that while NADPH oxidase exerts a strong bactericidal effect in the early stages of the infection, its effect becomes mainly bacteriostatic after 6 hours. This change in NADPH effects could be mediated by a transition in the population composition of *Salmonellae*-harbouring host cell types from macrophages to neutrophils, thus opening new avenues to explore the interactions between bacteria and different host immune cell types in detail. Similarly, in the enteric

phase of infection Maier *et al*. (Maier *et al.*, 2014) reported a transition from undetectable cidal activity to a strong bottleneck effect by day 2; they further suggested that the change could be attributable to NADPH oxidase and/or granulocyte activity.

Finally, using $ccr7^{-/-}$ mice, Kaiser *et al.* (Kaiser *et al.*, 2013) mapped bacterial migration to dendritic cell movement, while showing that bacterial migration from the gut to the caecal lymph node (cLN) is controlled by toxic reactive oxygen species using $cybb^{-/-}$ mice.

1.4.3.3. Characterisation of bacterial dynamics following immunisation

As highlighted in section 1.1.3, development of safe vaccines with durably protective immunogenicity and a cost-effective economic profile limit the repertoire of antigens, delivery systems and adjuvants that can be effectively used, and require the prediction of the immune responses elicited by the vaccine (Levine and Sztein, 2004). Alongside these considerations, the case of *Salmonella* being a facultative intracellular pathogen with a dual lifestyle is particularly challenging in terms of vaccine development, as protection against the pathogen requires both humoral and cellular immunity to eliminate both extracellular and intracellular bacteria respectively (Mastroeni, Villarreal-Ramos, and Hormaeche, 1993). To predict the immune response following immunisation, detailed knowledge about the host-pathogen interactions is needed (Rueckert and Guzman, 2012; Curtiss R. III, 2002; MacLennan, Martina and Micoli, 2014).

It has long been known that both live-attenuated (LV) and whole-cell killed (KV) vaccines increase the rate of blood clearance following an intravenous secondary challenge and reduce the bacterial load in the liver and spleen in the early stages of the infection (Collins, Mackaness and Blanden, 1966). However, whether these vaccines exert cidal or static effects and how their mode of action differs was not further investigated until 2014, when Coward *et al.* (Coward *et al*., 2014) combined ITS-based data with inferential modelling to conclude that the reduction in total numbers of viable bacteria observed in both the KV- and LV- groups is attributable to bactericidal effects in the first 24 hours post-challenge. However, the two vaccines were found to exert differential effects on the blood bacterial dynamics prior to colonisation of systemic tissues; while a large proportion of the inoculum was inactivated in the first 30’ post-challenge in the LV-group, no such inactivating effect was present in the KV-group.

Finally, the intra-organ dynamics were also predicted to be different for the two groups; bacteria in the LV-mice were predicted to be subject to higher bactericidal activity in the spleen, while bacteria in the KV-mice were subject to higher bactericidal activity in the liver. Unfortunately, this analysis was limited to simple models applied to the early stages of infection due to computational constraints when using a maximum-likelihood inference framework. As LVs are known to stimulate a cellular CD4+ and CD8+ T-cell response (Mastroeni, Villarreal-Ramos, and Hormaeche, 1992; Mastroeni, Villarreal-Ramos, and Hormaeche, 1993), further immunisation-challenge experiments were carried out on mice treated anti-$CD4^+$ and anti-$CD8^+$ antibodies to abrogate the cellular component of the immune response. The dynamics in this group were estimated to be similar to the baseline LV-immunised group for the first 72 hours of infection, leading to the conclusion that the induction of cellular immunity is responsible for neither the enhanced blood clearance of the inoculum nor the higher bactericidal activity in the spleen.

In the same year, Lim *et al.* (Lim *et al.*, 2014) independently studied the effects of a LV on the colonization of enteric, lymphoid and systemic organs following a secondary oral challenge with *S.* Typhimurium. Their findings on the effects of the live-attenuated vaccine on the systemic organs are compatible with those by Coward *et al.* (Coward *et al.*, 2014). Both groups observed that the spleens of LV-immunised mice were colonized by fewer bacteria, while the dissimilar ITS composition between the liver and spleen meant that the vaccine prevented free communication between the two organs. Furthermore, in the enteric phase of the infection, Lim *et al.* showed that LV-immunisation reduces the colonization of Payers' patches (PP) but does not eliminate it entirely. However, given the clear separation between the systemic and enteric phase of the infection, the vaccine-induced protection of the PPs does not noticeably affect the colonization of the spleen and liver and may, thus, be of little clinical importance.

Finally, Diard *et al.* (Diard *et al.*, 2019) evaluated the efficacy of an inactivated whole-cell oral vaccine against *S.* Typhimurium to show that while it induces high IgA titres, it fails to provide strong protection to the entire mouse cohort. This phenomenon was attributed to selection pressure in favour of variants that can escape IgA recognition, and was addressed by complementing the original vaccine formulation with a repertoire of surface antigens that generate IgA responses against these escape mutants.

1.4.3.4. Characterisation of bacterial dynamics during and following antibiotic exposure

At the level of the bacterial population, two studies have used the ITS technique to segregate between bacterial replication, death and migration following antibiotic exposure. Coupling the ITS technique with fluorescence microscopy and pharmacokinetic analysis, Kaiser *et al.* (Kaiser *et al.,* 2014) showed that *S.* Typhimurium survive ciprofloxacin treatment in the cLN, liver and spleen, and that the surviving bacteria were only killed with an antibiotic concentration ≥ 50-fold upper of the minimum inhibitory concentration (MIC) verifying their persister status.

Similarly, Rossi *et al.* (Rossi *et al.,* 2017) showed that a proportion of wild-type bacteria survive both ampicillin and ciprofloxacin treatment in the mesenteric lymph node (mLN), liver and spleen, and that the bacterial composition in the mLN is distinct from that of other organs, remaining largely unchanged throughout the antibiotic treatment. Furthermore, by comparing the effects of ampicillin and ciprofloxacin between murine infections with either a slow-growing or a wild-type *Salmonella* strain, they verified the positive correlation between bacterial growth rate and antimicrobial effect.

Other studies using fluorescence activated cell sorting (FACS) and fluorescent markers investigated bacterial replication and killing during antibiotic treatment in vivo, with single-bacterium resolution mainly in splenic homogenates. Claudi *et al.* (Claudi *et al.*, 2014) used TIMER, a fluorescent protein that changes colour over time, to quantify the proportions of rapidly-, moderately- and slowly-dividing bacteria in the spleens of naïve mice and mice exposed to ciprofloxacin. The reported a positive correlation between pre-treatment single-bacterium replication rates and the rate of clearance during treatment. However, given the differential prevalence of rapidly-, moderately- and slowly-growing bacteria in decreasing proportions, most bacteria surviving treatment originate from moderately-growing progenitors, as the absolute number of slow-replicating bacteria was low pre-treatment.

In contrast, in a murine model of early-onset, high-dose enrofloxacin treatment, Helaine *et al.* (Helaine *et al.*, 2014) concluded that surviving bacteria originate from a non-dividing bacterial subpopulation. Claudi *et al.* (Claudi *et al.*, 2014) reconciled this finding with their own conclusions by pointing out that in the experimental setup by Helaine *et*

al. the antibiotic treatment was started very early, before the inoculated bacteria adapted to the host, and as a result, the prevalence of non-dividing bacteria was exceptionally high.

Finally, Kanvatirth *et al.* (Kanvatirth *et al.*, 2020) used fluorescent-tagged antibodies against cell-type-specific markers to investigate the role of the difference in microenvironment between the range of host cell type occupied by *Salmonellae* during and after antibiotic treatment. They used ciprofloxacin, an antibiotic with good cellular penetration to rule out the hypothesis that surviving bacteria were not exposed to the antimicrobial agent. Furthermore, they reported a shift in the proportion of host cell types occupied by *Salmonellae* pre- and post-treatment, with neutrophils and dendritic cells being the predominant host cell type occupied by the bacteria surviving antibiotic treatment in the murine spleen.

1.4.3.5. Characterisation of the within-host dynamics of *in vivo* adapted *Salmonellae*

A large proportion of insights regarding pathogen behaviour comes from either *in vitro* studies, which are typically easier and more economical to perform, or *in vivo* studies in which animal hosts are inoculated with bacteria grown in lab cultures to ensure homogeneity in initial conditions and minimise the use of live hosts. However, it is becoming increasingly clear that even *in vivo* studies using *in vitro* grown bacteria may not reflect naturally acquired infections transmitted from host to host or acquired through a peripheral route and become systemic over time. The latter applies to *Salmonella* infections which, in human populations, are typically acquired through the faeco-oral route and can become systemic over time in immunocompromised individuals, young children and older populations; this suggests that there are two time windows when *Salmonellae* can develop host adaptations: a) between shedding from one host and uptake by the next, b) during the transition from the enteric to the systemic phase in orally acquired infections. Most *in vivo* studies studying the enteric phase of the infection use *in vitro* grown bacteria and those studying the systemic phase of the infection use intravenous inocula, bypassing the host adaptations that take place during the enteric phase that precedes systemic dissemination in most naturally acquired infections.

In 2015, Dybowski *et al.* (Dybowski *et al.*, 2015) used an ITS-based approach coupled with mathematical modelling to characterise the nature of *in vivo* adaptations that *S.* Typhimurium manifests upon systemic infection of a new host, compared to *in vitro* grown bacteria which are typically used in these studies. The distribution of ITS coupled with model-based inferences showed that *in vivo* adapted bacteria are better at surviving the initial bottleneck, as they enter a state of lower replication that makes them less sensitive to early bactericidal effects, and start their systemic spread up to 24 hours earlier than bacteria grown in culture.

1.4.3.6. Characterisation of host cell-bacterial interactions

Upon infection of the host, the progression of a bacterial infection is determined by bacterial virulence factors and host immune factors, as well as by the local host-pathogen interactions which stochastically depend on the microenvironment each bacterium encounters. For bacteria with both an intracellular and extracellular lifestyle, the effect of these stochastic host-pathogen interactions is amplified, as their proliferation or extinction becomes multifactorially dependent on: 1) local concentration gradients of humoural elements of the host's immunity, 2) the local distribution of host cell types and 3) the interaction of the two that often determines the phenotype of the present host cells (e.g. cytokine-activated versus resting macrophages). Furthermore, the intracellular life mode of a bacterium can range from high intracellular density to sparse host cell occupancy; in the latter case, bacteria continuously escape infected host cells and invade new, adding further layers of stochasticity.

While earlier *in vitro* experiments with macrophage cultures had shown that *S.* Typhimurium grows to high intracellular densities (Harrington and Hormaeche, 1986), *ex vivo* fluorescence microscopy on single liver macrophages showed that the majority of macrophages harbour 1-2 bacteria, with few macrophages harbouring higher bacterial loads (4-6 bacteria), and that the increase in total bacterial population size is paralleled by an increase in infected macrophages (Sheppard *et al.*, 2013). A later modelling study considered a range of possible biological mechanisms to explain the heterogeneity in bacterial numbers per macrophage, concluding that the model with the strongest support was one with homogeneous host cell permissiveness to bacterial proliferation, where host cell burst events occur stochastically as a result of uniform bacterial division rates,

bacterial accumulation in the intracellular environment and host cell lysis (Brown *et al.*, 2006).

While the studies by Sheppard *et al.* and Brown *et al.* focused on macrophages, it is well-established that *Salmonellae* interact with a range of host cell types including dendritic cells and neutrophils (reviewed in Castanheira and Portillo, 2017). The relative abundance of these cell types fluctuates over the course of the infection, as well as during antibiotic treatment or in immunised hosts. Kanvatirth *et al.* (Kanvatirth *et al.*, 2020) studied the relationship between the relative abundance of host immunity cell types and the numbers of viable bacteria in the spleen of mice during and post-antibiotic treatment with ciprofloxacin. They reported that dendritic cells, macrophages and neutrophils are the main host cell types associated with *Salmonellae* during and following ciprofloxacin exposure. Depleting neutrophils during treatment resulted in an increase in the number of viable *Salmonellae* at the end of antibiotic treatment, while the opposite effect was observed when dendritic cells and macrophages are depleted. Taken together, these results reveal a dichotomy in the role of host cell types in clearance and proliferation.

Table 1.1: A thematic summary of the experimental studies investigating the within-host dynamics of bacterial infections, adapted and extended from Vlazaki, Huber and Restif, 2019

Thematic Area	Bacterial Identity	Experimental Technique	Characterised Process	Study
Characterisation of colonisation bottlenecks & intra-organ growth and inter-tissue migration	*Salmonella enterica* serovar *Typhimurium*	Plasmid segregation	Replication and death	Gulig and Doyle, 1993
		ITS	Replication, migration and death	Grant *et al.*, 2008 Kaiser *et al.*, 2013 Lam and Monack, 2014
	Listeria monocytogenes	Plasmid segregation	Replication and death	Bakardjiev *et al*., 2006
		ITS	Replication, migration and death	Melton-Witt *et al.*, 2012 Zhang *et al.*, 2017
	Uropathogenic *Eschericia coli*	ITS	Replication, migration and death	Schwartz *et al*., 2011 Walters *et al.*, 2012
	Eschericia coli K12	Plasmid segregation	Replication and death	Meynell, 1959 Mow and Meynell, 1968
	Bacillus anthracis	ITS	Replication, migration and death	Lowe *et al*., 2013
		Multiple fluorescent strains	Colonisation	Plaut *et al*., 2012
	Streptococcus pneumoniae	ITS	Replication, migration and death	Maier *et al.*, 2014 Gerlini *et al.*, 2014
	Borrelia burgdorferi	ITS	Replication, migration and death	Rego *et al*., 2014
	Vibrio cholerae	ITS	Replication, migration and death	Abel *et al*., 2015
		Multiple fluorescent strains	Colonisation	Millet *et al*., 2014
	Yersinia pestis	ITS	Replication, migration and death	Gonzalez *et al*., 2015
	Mycobacterium tuberculosis	ITS	Replication, migration and death	Martin *et al*., 2017
	Haemophilus influenzae b	Plasmid segregation	Replication and death	Moxon and Murphy, 1978
Identification of host immunity mechanisms responsible for infection control	*Salmonella enterica* serovar *Typhimurium*	Plasmid segregation	Replication and death	Hormaeche, 1980
		ITS	Replication, migration and death	Grant *et al*., 2008 Kaiser *et al*., 2013 Maier *et al*., 2014
	Vibrio cholerae	Multiple fluorescent strains	Colonisation	Millet *et al*., 2014
Characterisation of bacterial dynamics following immunisation	*Salmonella enterica* serovar *Typhimurium*	ITS	Replication, migration and death	Coward *et al*., 2014 Lim *et al*., 2014
Characterisation of bacterial dynamics in the gut microbiome	Multi-species bacterial samples	Fluorescent marker	Replication	Myhrvold *et al.*, 2015
		Peak-to-trough ratio		Korem *et al*., 2016
	Salmonella enterica serovar *Typhimurium*	ITS	Colonisation, growth	Wotzka *et al*., 2019

Characterisation of bacterial dynamics during or post- antibiotic administration	*Salmonella enterica* serovar *Typhimurium*	ITS	Replication, migration and death	Kaiser *et al.*, 2014 Rossi *et al.*, 2017 Vlazaki *et al.*, 2020
		Fluorescent marker	Colonisation	Diard *et al.*, 2014
			Replication	Claudi *et al.*, 2014 Helaine *et al.*, 2010 Helaine *et al.*, 2014
			Plasmid conjugation	Bakkeren *et al.*, 2019
			Chronic bacterial carriage	Monack, Bouley and Falkow, 2004
	Eschericia coli	Fluorescent marker	Replication	Roostalu *et al.*, 2008
		Plasmid segregation	Replication, death	Frenoy and Bonhoeffer, 2018
	Staphylococcus aureus	Plasmid segregation	Replication, death	McVicker *et al.*, 2014
	Salmonella enterica serovar *Typhimurium*	Fluorescent marker	Intracellular dynamics	Kanvatirth *et al.*, 2020 Sheppard *et al.*, 2003 Brown *et al.*, 2006
Characterisation of bacterial adaptations following colonisation	*Staphylococcus aureus*	Multiple fluorescent strains	Colonisation	Prajsnar *et al.*, 2012
	Salmonella enterica serovar *Typhimurium*	ITS	Replication, death	Dybowksi *et al.*, 2015 Dybowksi *et al.*, 2017

1.5. Thesis Structure and Novel Contributions

1.5.1. Thesis Motivation

Large-scale mechanistic modelling using population-level data in microbiology is lacking. Using data-driven models in the field of population dynamics microbiology has been limited by the lack of efficient computational methods to render complex mathematical models feasible. Existing data-driven modelling studies are limited to: a) short time intervals, usually applied to data from the early phases of a bacterial infection (Grant *et al.*, 2008; Kaiser *et al.*, 2013; Coward *et al.*, 2014), b) using a single model without comparing the goodness-of-fit of rival models, c) models with simple dynamics and/or few compartments (Brown *et al.*, 2006). The motivation of the present thesis has been to approach the modelling-data interface pleiotropically, illustrating the range of applications of mathematical modelling, as well as the diversity of biological questions that can be addressed through data-driven models, summarised in Figure 1.4.

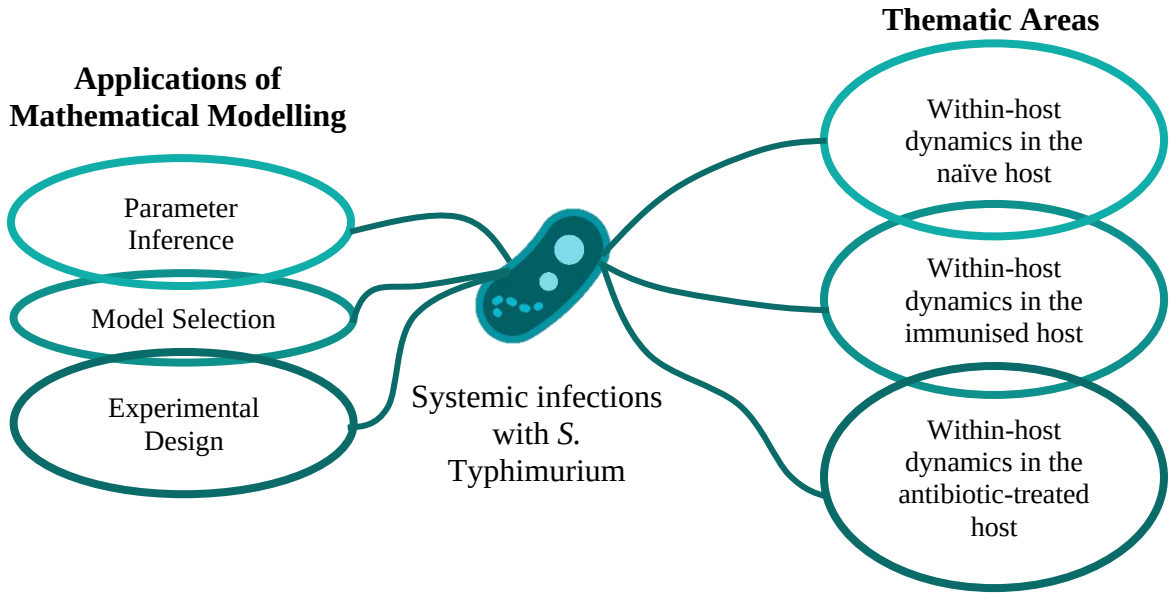


Figure 1.4: Schematic representation of the outline of the present thesis

In the present thesis I aim to:

1. provide the first application of a validated, computationally-efficient tool developed to allow large-scale modelling of population-level data from studies in within-host bacterial dynamics,

2. provide a series of case studies of large-scale, data-driven modelling studies covering the entire timeline of available experimental data, serving as blueprints for an era of more extensive and complete integration between experimental data and mathematical modelling,

3. illustrate the multiple uses of mathematical modelling when integrated with experimental data, including: parameter inference, model selection and experimental design, and

4. analyse previously published ITS-based datasets and generate biological insights that have helped address knowledge gaps with regards to the impact of antibiotic therapy and immunisation on the within-host dynamics of *S.* Typhimurium in the systemic phase of infection.

1.5.2. Thesis Structure

The present thesis includes 7 chapters, of which 4 chapters (3-6) correspond to results obtained during this PhD. Chapter 1 defines the methodological and current knowledge framework, chapter 2 details the methods used and chapter 7 provides the general discussion about the implications and future directions of my research. Their contents are briefly outlined below:

- **Chapter 3**: I use two published ITS-based datasets with bacterial counts from the blood, liver and spleen collected over 72 hours from systemic infections with *S.* Typhimurium in naïve mice (Coward *et al.*, 2014). I combine these data with stochastic compartmental models, assess their goodness-of-fit and perform **model selection**. Finally, I use the better-fitting model for **parameter inference** to quantitatively characterise the **dynamics in the naïve host.**

- **Chapter 4:** I use a published ITS-based dataset with bacterial counts from the blood, liver and spleen collected over 72 hours from systemic infections with *S.* Typhimurium in 3 groups of mice: naïve mice or mice previously immunised with either a live-attenuated or whole-cell killed vaccine (Coward *et al.*, 2014). I use the model identified in Chapter 3 for **parameter inference** to compare the **impact of** the two **vaccine formulations** on these dynamics.

- **Chapter 5**: I use a published ITS-based dataset with bacterial counts from the blood, liver, spleen and mesenteric lymph nodes, collected over 14 days from a systemic infection with *S.* Typhimurium in 2 groups of **antibiotic-treated mice**: ciprofloxacin- and ampicillin-treated (Rossi *et al.*, 2017). I formulate stochastic compartmental models with either a single or dual phenotype, assess their goodness-of-fit, perform model **selection** and make **inferences** about the effect of the two antibiotics on the within-host dynamics of *S.* Typhimurium. I address questions relevant to the **origins, migration patterns and post-treatment behaviour of bacterial populations that survive antibiotic treatment.**

- **Chapter 6**: I extend the likelihood-free inference approach previously developed by Price *et al.* (Price *et al.*, 2017) to **an experimental design tool** that can *a priori* quantify the quality of **inference precision** provided by experimental designs with different design factors. I define the determinant of the parameter variance-covariance matrix as a measure of global precision and use a simulation study to show excellent concordance with the width of 95% confidence intervals obtained by parametric bootstrap. I assess the effect of three design factors (number of replicates, choice of sampling time points, initial number of bacterial copies per ITS), and conclude with a series of case studies representing typical patterns of within-host dynamics typically observed in bacterial infections.

1.5.3. Novel Contributions

As noted in 1.5.1 and 1.5.2, the present thesis contains a mix of studies aimed at obtaining novel biological insights and studies oriented towards developing and validating computational method for use with ITS-based data. The novel contributions made can be summarised as follows:

Chapter 3: *S.* Typhimurium dynamics in the naïve host

1. In the naïve murine host, inter-tissue mixing occurs during the first 6 hours post-inoculation, earlier than previously thought.
2. In the naïve host, bacteria migrate asymmetrically to the liver and spleen; bacteria preferentially migrate to the liver during the first 6 hours post-inoculation.
3. With parameter estimates for inter-tissue bacterial migration and intra-organ death and replication, the systemic phase of salmonellosis can be broken down to 3 distinct phases starting from inoculation: a) 6 hours of niche establishment in the organs and intra-organ killing, followed by b) 18 hours of exponential growth, and c) 48 hours of stable, marginal intra-organ growth.

Chapter 4: S. Typhimurium dynamics in the immunised host

1. The live-attenuated and whole-cell killed vaccines exert different effects on the within-host bacterial dynamics. While the former acts on the liver, the latter acts on the spleen.
2. The inferred effects of the live-attenuated vaccine are consistent with earlier studies showing that it induces higher levels of IgG2 antibody compared to the killed-vaccine.

Chapter 5: *S.* Typhimurium dynamics in the antibiotic-treated host

1. The dynamics in the ciprofloxacin-treated mice can be explained by a single-phenotype model, while those in the ampicillin-treated mice require a more complex dual-phenotype model including a non-growing, antibiotic-recalcitrant bacterial subpopulation.

2. In ampicillin-treated mice, the non-growing, antibiotic-recalcitrant subpopulations emerge after 48 hours of treatment in the spleen and mesenteric lymph nodes. My model also infers spleen-to-liver bacterial transfer.

3. Unlike ciprofloxacin, ampicillin exerts bacteriostatic effects in the blood and in the late phase of treatment in the spleen and mesenteric lymph node.

4. During relapse following antibiotic treatment, the previously non-growing populations reconstitute the bacterial population in the mesenteric lymph node and spleen, where the growing population went extinct prior to antibiotic withdrawal, but also in the liver as they replicate faster than the growing population.

Chapter 6: A novel tool for experimental design in ITS-based experiments

1. The determinant of the parameter variance-covariance matrix obtained by parametric bootstrap represents a reliable summary statistic for inference precision, and can used in optimal experimental design in a likelihood-free inference framework.

2. In ITS-based studies, a higher number of biological replicates, a higher number of initial copies per ITS and a selection of sampling timepoints based on the pattern of observed dynamics all lead to higher-precision estimates.

3. With regards to optimal resource allocation and animal welfare, using experimental infections with a lower number of ITS per host at higher number of copies per ITS yields higher-precision inference than a higher number of ITS per host at a lower number of copies per ITS.

2.

Experimental Data & Mathematical Methods

Summary Points:

- ITS-based datasets consist of observations of copies per ITS in each tissue of interest at different sampling timepoints.
- ITS-based data can be combined with compartmental models, with each compartment representing the bacterial subpopulation residing in each tissue of interest at time t.
- The master or forward Kolmogorov equation describes the probability of observing a given number of copies per ITS in each compartment at time t. For complex biological systems, the master equation can only be numerically approximated; this is computationally intensive and time-consuming.
- An alternative to solving the master equation is summarising the distribution of bacterial copies per ITS in each compartment by a set of summary statistics: the first- and second- order moments.
- Parameter inference is performed by minimising the Kullback-Leibler divergence between the experimental and calculated moments for a range of parameter values.
- Assessment of model goodness-of-fit and quantification of inference precision both require a parametric bootstrap approach.
- Experimental design in terms of optimising inference precision is performed by minimising the determinant of the parameter estimate variance-covariance matrix obtained by parametric bootstrap.

Declaration of Originality

This chapter includes written material that has been verbatim reproduced from a paper currently in press (Vlazaki, Price and Restif, 2020), for which I certify that I am the sole contributor in terms of study design, execution and manuscript writing. I also certify that I have included the relevant citations, where appropriate.

2.1. The ITS technique: experimental output and dataset structure

ITS represent a large class of marker-based techniques developed over the last decade to study the demographic processes that collectively determine the total observable microbial load in a tissue of interest (Grant *et al*., 2008). Libraries of isogenic tagged strains are generated by uniquely modifying the non-coding portion of the bacterial genome with distinctive, short nucleotide sequences; the resulting ITS are genetically distinguishable, yet phenotypically identical. The ITS are typically mixed in equiproportional inocula and administered to the animal host. Animals are culled at predetermined time points to harvest or partly sample tissues of interest, and a total bacterial count is obtained per tissue (Grant *et al*., 2008; Abel *et al*., 2015). The samples are processed to determine the ITS composition in each tissue, originally by qPCR (e.g., Grant *et al*., 2008; Kaiser *et al*., 2013; Kaiser *et al*., 2014) and more recently by next-generation sequencing (e.g., Rossi *et al*., 2017; Lim *et al*., 2014). The final structure of the dataset consists of number of copies per ITS in each tissue of interest per time point, as summarised in Figure 2.1.

By obtaining snapshots of spatiotemporal data recapitulating the changing ITS composition of bacteria in different tissues, it is possible to characterise the unobserved processes that underpin the *in vivo* progression of a bacterial infection. Supplementing these data with compartmental mechanistic models, it is possible to quantify the rates at which these unobserved dynamical processes progress with time. A unique feature of ITS-based studies is the multiplicity of infections with different tagged strains within each host organism. Under the assumptions of independent action for bacteria and homogeneity in dynamics both between ITS and between host organisms, each ITS in the inoculum can be treated as an independent infection, with the total number of biological replicates in a dataset being equal to the number of ITS per host multiplied by the number of hosts. Experiments produce measurements of the number of copies of each ITS in anatomical compartments of interest at the time that each host animal is killed, and inference about the within-host dynamics "fills in" the gap between the sampled timepoints.

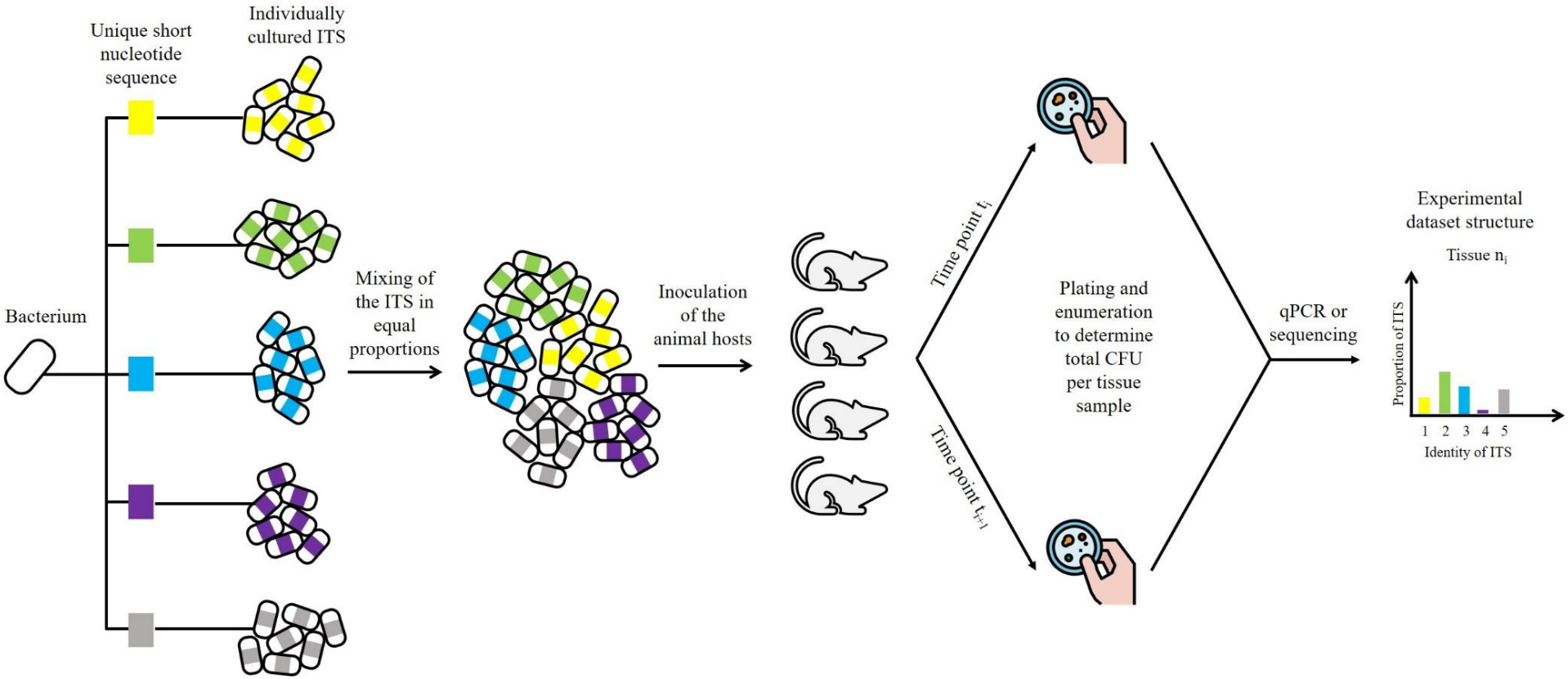


Figure 2.1: Summary illustration of the ITS technique and the structure of the experimental dataset

2.2. Conceptual Model Structure

Mechanistic models applied to ITS data typically have a compartmental structure, with each compartment representing a tissue of interest, and model parameters corresponding to the rates at which the unobserved biological processes change with time. Figure 2.2 illustrates an example of a model with three compartments, all possible transitions between them and intra-compartmental bacterial death and replication.

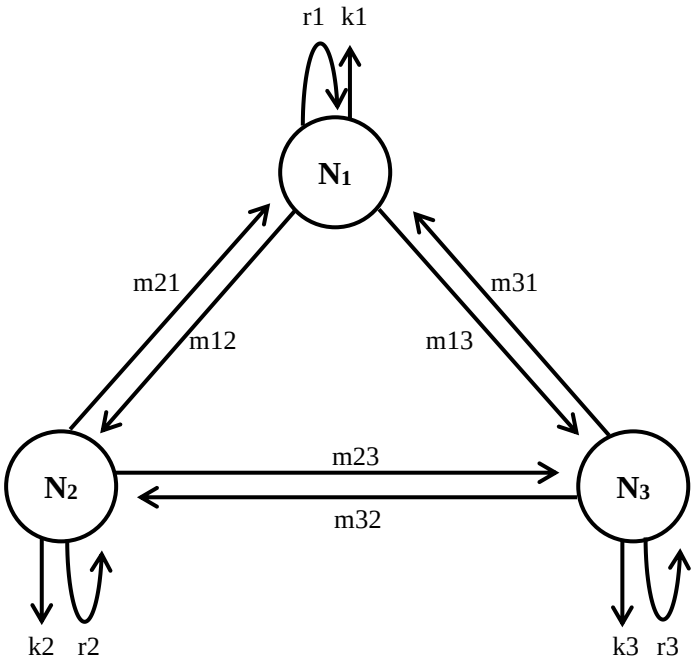


Figure 2.2: Conceptual structure of a mechanistic model with three compartments $\{N_1, N_2, N_3\}$, intra-organ bacterial killing (k_i) and replication (r_i), and inter-tissue migration (m_{ij}).

Earlier experimental studies on the pathogenesis of systemic salmonellosis in the murine host have revealed that upon intravenous inoculation, the liver and spleen act as major bacterial reservoirs with bloodstream-mediated bacterial migration (Grant *et al.*, 2008), while for infections obtained through the orogastric route, the mesenteric lymph node is an additional bacterial reservoir (Rossi *et al*., 2017; Grant *et al*., 2008; Kaiser *et al*., 2013). This biological insight motivated compartmental models encompassing the blood, liver, spleen and mesenteric lymph node (Grant *et al*., 2008; Coward *et al*., 2014). To enable parameter inference of both intra-organ (e.g. killing) and inter-organ (migration) processes, a stochastic model was used in which the number of bacteria per ITS in each organ was treated as a random variable. The first- and second- order moments were used to characterise this distribution and the system was modelled as a multivariate Markov process involving independent stochastic variables corresponding to individual tissues.

2.3. Mathematical Model Structure

[reproduced verbatim from Vlazaki, Price and Restif, 2020; Supplementary Methods]

Multivariate Markov processes are defined as processes involving a number of independent stochastic variables, whose future state only depends on the present state and not the past (Markov, 1906). The dynamics of such processes collectively determine the probability that the system is at a given state at any particular time point; this probability can be summarised by the master equation (2.3.1). In bacterial dynamics, the state of the system refers to a vector consisting of the number of bacteria residing in the tissues of interest.

For a multivariate process, the master equation becomes impossible to solve analytically. Numerical approximations of the master equation are inaccurate and time-consuming to obtain. To address this challenge, Coward *et al.* (Coward *et al.*, 2014) and Price *et al.* (Price *et al.*, 2017) showed that under the assumption of constant rates for the within-host dynamics in each time interval (first-order dynamics) and the simplification that ITS are normally distributed in each tissue, first- and second- moments can be used as sufficient summary statistics to describe the state that the system is in. It is, then, possible to derive mathematical equations for the first- (2.3.2) and second-order moments (2.3.3) of the experimentally observed states by adapting a moment-closure method for stochastic models in population biology described by Singh and Hespanha (Singh and Hespanha, 2006). Given the initial first- (means) and second-order moments (variances, covariances) per organ per time point and values for the parameters governing the dynamics of the system, the equations describing the evolution of the moments can be solved to obtain the moments at a later timepoint, thus bypassing the need to numerically evaluate the master equation.

Let $M_1(t)$ be a vector of the first-order moments for the three-compartment system in Figure 2.1, and $M_2(t)$ be the vector of second-order moments (variances, covariances) in the same system. $M_1(0)$ and $M_2(0)$ refer to the first- and second- moments at t=0 respectively. To calculate the moments at time t, $M_1(t)$ and $M_2(t)$, it is possible to solve equations 2.3.4 and 2.3.5 numerically in R using the SPEEDI.R package (Price *et al.*, 2017).

Equation 2.3.1: The master equation for the three-compartmental model

*Interpreted as the instantaneous probability that at time **t** there are n_1 bacteria in the 1st compartment, n_2 bacteria in the 2nd compartment and n_3 bacteria in the 3rd compartment*

The rate m_{21} is multiplied by the probability of a migration event from the 2nd to the 1st compartment. Each bacterium in the population has the same probability of migrating, hence the multiplication with the total number of bacteria ($n_2 + 1$) in the 2nd compartment, prior to the migration event

$$\dot{P}(n_1, n_2, n_3) = m_{12}P(n_1 + 1, n_2 - 1, n_3)(n_1 + 1) + m_{13}P(n_1 + 1, n_2, n_3 - 1)(n_1 + 1) + m_{21}P(n_1 - 1, n_2 + 1, n_3)(n_2 + 1)$$
$$+ m_{31}P(n_1 - 1, n_2, n_3 + 1)(n_3 + 1) + m_{23}P(n_1, n_2 + 1, n_3 - 1)(n_2 + 1)$$
$$+ m_{32}P(n_1, n_2 - 1, n_3 + 1)(n_3 + 1) + r_1P(n_1 - 1, n_2, n_3)(n_1 - 1) + r_2P(n_1, n_2 - 1, n_3)(n_2 - 1)$$
$$+ r_3P(n_1, n_2, n_3 - 1)(n_3 - 1) + k_1P(n_1 + 1, n_2, n_3)(n_1 + 1) + k_2P(n_1, n_2 + 1, n_3)(n_2 + 1) + k_3P(n_1, n_2, n_3 + 1)(n_3 + 1)$$
$$- [(r_1 + k_1 + m_{12} + m_{13})n_1 + (r_2 + k_2 + m_{21} + m_{23})n_2 + (r_3 + k_3 + m_{31} + m_{32})n_3]P(n_1, n_2, n_3)$$

where $\dot{P}(n_1, n_2, n_3)$ is an abbreviated notation for $\dot{P}(n_1(t), n_2(t), n_3(t))$ and represents a system of ODEs tracking the probabilities $P(n_1(t), n_2(t), n_3(t))$.

It is possible to approximate the dynamics of multivariate Markovian processes using moment-closure methods to derive a system of ODEs for the moments of the state variables from the master equation (2.3.1) of any stochastic model (Price *et al.*, 2017). By considering the first- and second-order moments as sufficient summary statistics, it is possible to derive a closed, small-dimension system allowing fast numerical solution of these moments at any time point. From the general equation for the j^{th} moment of a distribution described by Singh and Hespanha (Singh and Hespanha, 2006), it is possible to obtain a set of differential equations for the first-order (Equation 2.3.2) and second-order (Equation 2.3.3) moments as a function of time.

$$\frac{\partial M_1(t)}{\partial t} = A \times M_1(t) \quad \textbf{(2.3.2)}$$

$$\frac{\partial M_2(t)}{\partial t} = C \times M_2(t) + B \times \exp(tA) \times M_1(0) \quad \textbf{(2.3.3)}$$

Price *et al*. (Price *et al*., 2017) provides the matrix form solution to this set of differential equations, shown in 2.3.4 and 2.3.5 for the first- and second-order moments respectively.

$$M_1(t) = \exp(At) \times M_1(0) \quad \textbf{(2.3.4)}$$

$$M_2(t) = \exp(tC) \times M_2(0) + \exp(tC) \times \int_0^t [\exp(-sC) \times B \times \exp(sA)\, ds] \times M_1(0) \quad \textbf{(2.3.5)}$$

$M_1(t)$ and $M_2(t)$ represent the first- and second-order moments in vector form respectively, as shown in 2.3.6 and 2.3.7.

$$M_1(t) = \begin{pmatrix} E[N_1(t)] \\ E[N_2(t)] \\ E[N_3(t)] \end{pmatrix} \quad \textbf{(2.3.6)}$$

$$M_2(t) = \begin{pmatrix} Var(N_1(t)) \\ Var(N_2(t)) \\ Var(N_3(t)) \\ Var(N_1(t), N_2(t)) \\ Var(N_1(t), N_3(t)) \\ Var(N_2(t), N_3(t)) \end{pmatrix} \quad \textbf{(2.3.7)}$$

The matrices A, B and C appearing in equations 2.3.2 – 2.3.5 are populated with the model parameters. For a three-compartment model with all allowed transitions between compartments illustrated as an example in 2.2, matrices A, B and C take the form below (2.3.8 – 2.3.11).

$$A = \begin{pmatrix} r_1 - k_1 - m_{12} - m_{13} & m_{21} & m_{31} \\ m_{12} & r_2 - k_2 - m_{21} - m_{23} & m_{32} \\ m_{13} & m_{23} & r_3 - k_3 - m_{31} - m_{32} \end{pmatrix} \quad \textbf{(2.3.8)}$$

$$B = \begin{pmatrix} r_1 + k_1 + m_{12} + m_{13} & m_{21} & m_{31} \\ m_{12} & r_2 + k_2 + m_{21} + m_{23} & m_{32} \\ m_{13} & m_{23} & r_3 + k_3 + m_{31} + m_{32} \\ -m_{12} & -m_{21} & 0 \\ -m_{13} & 0 & -m_{31} \\ 0 & -m_{23} & -m_{32} \end{pmatrix} \quad \textbf{(2.3.9)}$$

$Let\ C = [D|E]$, C being the concatenation of matrices D and E.

$$D = \begin{pmatrix} 2(r_1 - k_1 - m_{12} - m_{13}) & 0 & 0 \\ 0 & 2(r_2 - k_2 - m_{21} - m_{23}) & 0 \\ 0 & 0 & 2(r_3 - k_3 - m_{31} - m_{32}) \\ -m_{12} & -m_{21} & 0 \\ -m_{13} & 0 & -m_{31} \\ 0 & -m_{23} & -m_{32} \end{pmatrix} \quad \textbf{(2.3.10)}$$

$$E = \begin{pmatrix} 2m_{21} & 2m_{31} & 0 \\ 2m_{12} & 0 & 2m_{32} \\ 0 & 2m_{13} & 2m_{23} \\ (r_1 + r_2 - k_1 - k_2 - m_{12} - m_{13} - m_{21} - m_{23}) & m_{21} + m_{23} & m_{31} + m_{32} \\ m_{12} + m_{13} & (r_1 + r_3 - k_1 - k_3 - m_{12} - m_{13} - m_{31} - m_{32}) & m_{32} + m_{31} \\ m_{13} + m_{12} & m_{23} + m_{21} & (r_2 + r_3 - k_2 - k_3 - m_{21} - m_{23} - m_{31} - m_{32} \end{pmatrix} \quad \textbf{(2.3.11)}$$

2.4. The Kullback-Leibler divergence as a cost function

In all inference frameworks used in data-driven mathematical models, parameter inference is achieved by comparing the output of a mathematical model with the available experimental observations, and choosing the parameter set that makes the two as identical as possible. Cost or utility functions are used to measure how different or similar these quantities are; optimising these functions identifies the estimates for the unknown parameters. Figure 2.3 illustrates the example of ITS-based experiment to showcase the inference process in the likelihood-free framework and highlight the role of the cost function.

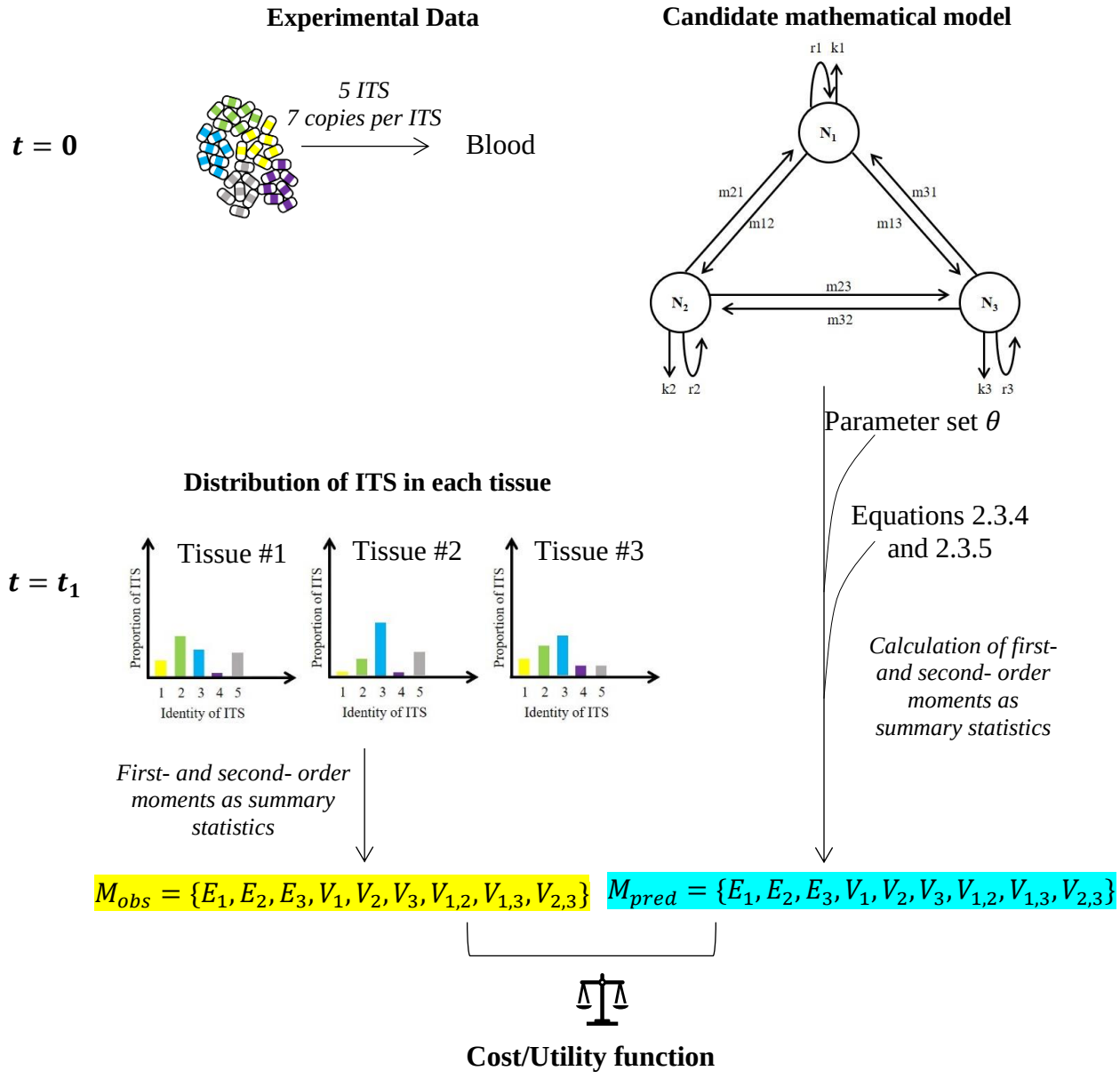


Figure 2.3: Schematic representation of the inference process and the role of the cost/utility functions as one-dimensional measures of similarity between model and experimental output

To quantify the difference between the experimentally observed and model-predicted moments in the likelihood-free inference framework, it is possible to use a range of divergence measures including the Chi-squared distance, the Mahalanobis distance, the Hellinger distance, the Kullback-Leibler (KL) divergence and others. Price *et al*. (Price *et al*., 2017) compared their performance by evaluating the mean absolute relative error of parameter estimates in a three-compartment model. As shown in Figure 2.4, the performance of the Kullback-Leibler divergence is superior, with mean absolute relative errors consistently at ~ 0.15 (15%) across 30 simulated scenarios.

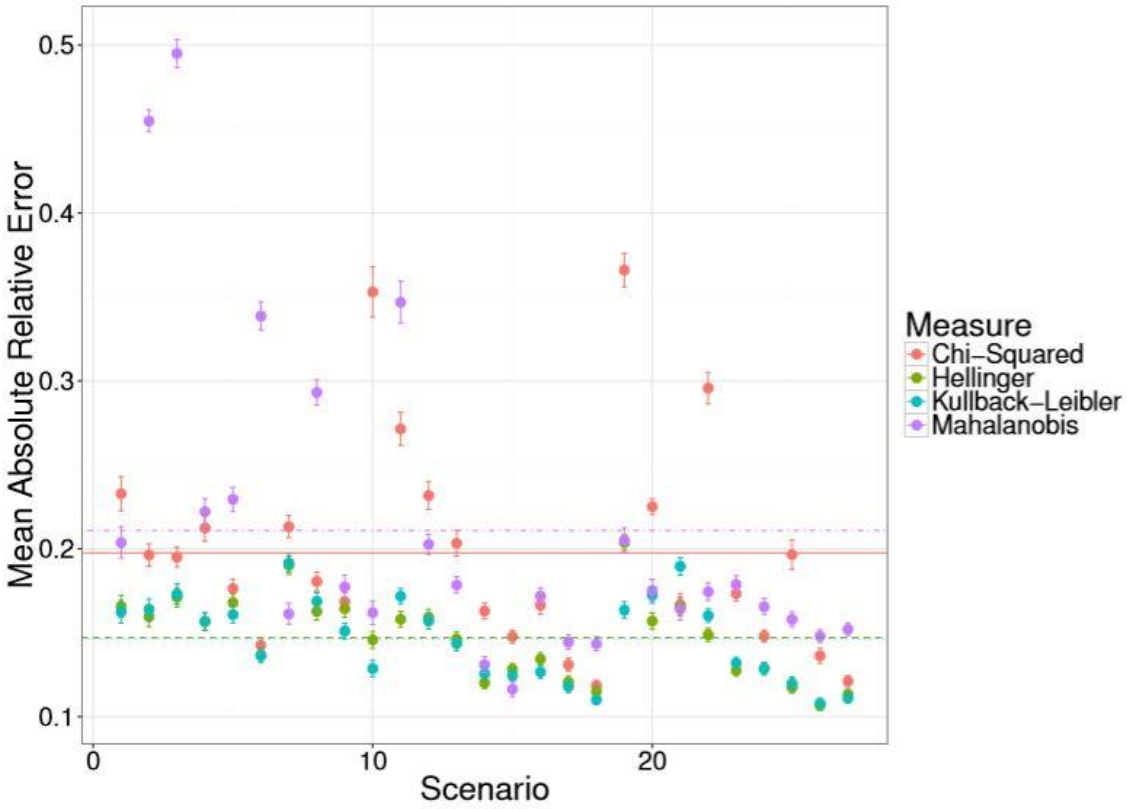


Figure 2.4: Comparative performance of different cost functions
Plot of mean absolute relative errors of parameter estimates for a three-compartment model, for a range of divergence measures across 30 scenarios, averaged across each of the 100 simulations at each observation time. The horizontal lines correspond to the average mean absolute relative error for each method *[reproduced from Price et al., 2017]*

In this thesis, I used a simplified expression for the KL-divergence as the preferred cost function. Assuming a true distribution *P* for a continuous random variable and an artificial distribution Q for the same random variable, the KL-divergence measures the information lost when using an artificial probability distribution Q (model-predicted distribution) to describe the true probability distribution *P* (experimentally observed distribution). For distributions P and Q of a continuous random variable, the KL divergence is defined by 2.4.1:

$$\Delta_K(P\|Q) = \int_{\mathbb{R}_d} p(x) log \frac{p(x)}{q(x)} dx \qquad \textbf{(2.4.1)}$$

To facilitate the numerical evaluation of this expression using only the first- and second-order moments of the distributions P and Q, Price *et al.* (Price *et al.*, 2017) made the simplifying assumption that for n ITS in the initial inoculum, the distribution of ITS copies in each compartment is a multivariate normal distribution with means $\mu = \{\mu_1, \mu_2, \mu_3 \dots, \mu_n\}$ and variances $V = \{V_1, V_2, V_3 \dots, V_n\}$, transforming equation 2.4.1 into 2.4.2, which performs well as a cost function in this moments-based inference framework (Vlazaki, Price and Restif, 2020). Subscripts *obs* and *pred* represent the observed and predicted vectors respectively.

$$\Delta_K = \frac{1}{2}\left[tr\left(V_{obs}^{-1}\ V_{pred}\right) - n + \mu^T V_{obs}^{-1}\,\mu + log\frac{|V_{obs}|}{|V_{pred}|}\right], where \qquad \textbf{(2.4.2)}$$

$$\mu = \mu_{pred} - \mu_{obs}$$

2.5. Linear regressions to correct experimental data for noise

In ITS-based experiments, raw experimental data consist of: 1) the total number of bacteria per tissue per time point and 2) the proportions at which each ITS is represented in the total bacterial count, as determined by either qPCR or next-generation sequencing. Accounting for the proportion of tissue sampled, it is possible to calculate the number of copies per ITS per tissue at each time point. However, the raw data are subject to at least two sources of error. First, there is sampling error due to incomplete tissue sampling which is typically secondary to technical constraints during the experimental procedure (e.g. it is impossible to obtain 100% of the blood). Second, noise is introduced by the ITS detection technique when quantifying the true proportions of tagged strains in a minute sample used for detection.

As inference relies on the comparison of experimental data with noise-free simulated data, the two known sources of noise should be accounted for in the experimental data prior to inference.

Since the generating processes of both sources of noise are well-understood, both errors can be modelled accordingly, and any "perfect" observation can be converted to noisy through appropriate transformations using the model(s) for the source(s) of noise. The sampling error due to partial sampling of the tissue can be modelled as a binomial process (2.5.1). In terms of ITS-detection, the qPCR detection noise has been empirically modelled as a log-normal process by plating known numbers of colonies of each ITS, putting them through qPCR and fitting a family of models to characterise the relationship between qPCR reads and ITS colonies in the sample. Using a maximum likelihood approach, Coward *et al.* (Coward *et al.*, 2014) identified the best-fitting model as shown in 2.5.2.

$\boldsymbol{v \sim Binom(n, \varphi)}$, where:
v: the number of copies per ITS in the sample
n: the number of copies per ITS in the whole tissue **(2.5.1)**
φ: the proportion of tissue sampled

$\boldsymbol{\log(\Omega_v) \sim N(\log(v), 0.2674e^{-0.0148v})}$, where:
Ω_v: number of qPCR reads per ITS
v: the true number of copies per ITS in the sample **(2.5.2)**

For ITS data analysed using next-generation sequencing, it is possible to obtain a low count of reads even in the absence of ITS (yellow curve in Figure 2.5). To account for this, Rossi *et al.* (Rossi *et al.*, 2017) established a technical noise cut-off (green line in Figure 2.5) below which it is assumed that sequence reads do not reflect the true presence of ITS copies in the sample.

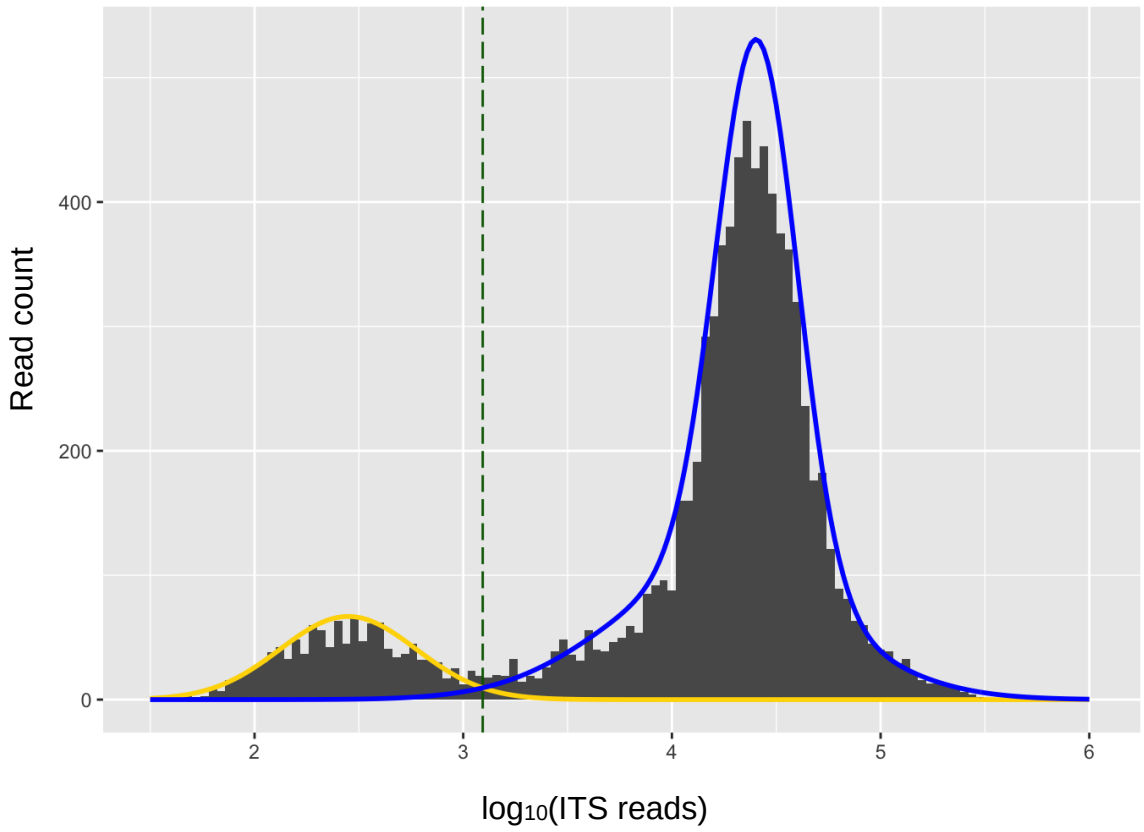


Figure 2.5: Technical cut-off to account for ITS detection noise when using next-generation sequencing *[reproduced from Rossi et al., 2017]*

Having characterised the noise processes, I used a reverse engineering approach to convert the observed noisy moments back to their true "perfect" state. To achieve this, I simulated large numbers of "perfect" datasets, transformed them into noisy data by applying the noise models, calculated the corresponding "perfect" and "noisy" moments, and used linear regression models weighted by the distance of the simulated data value to the experimental value to find an expression that describes their relationship. An example of this algorithm applied to simulated "perfect" and simulated noisy data from a three-compartment model is illustrated in Figure 2.6. I use this expression obtained from the linear regression model to convert my noisy experimental data into "perfect" data, which can be used for parameter inference.

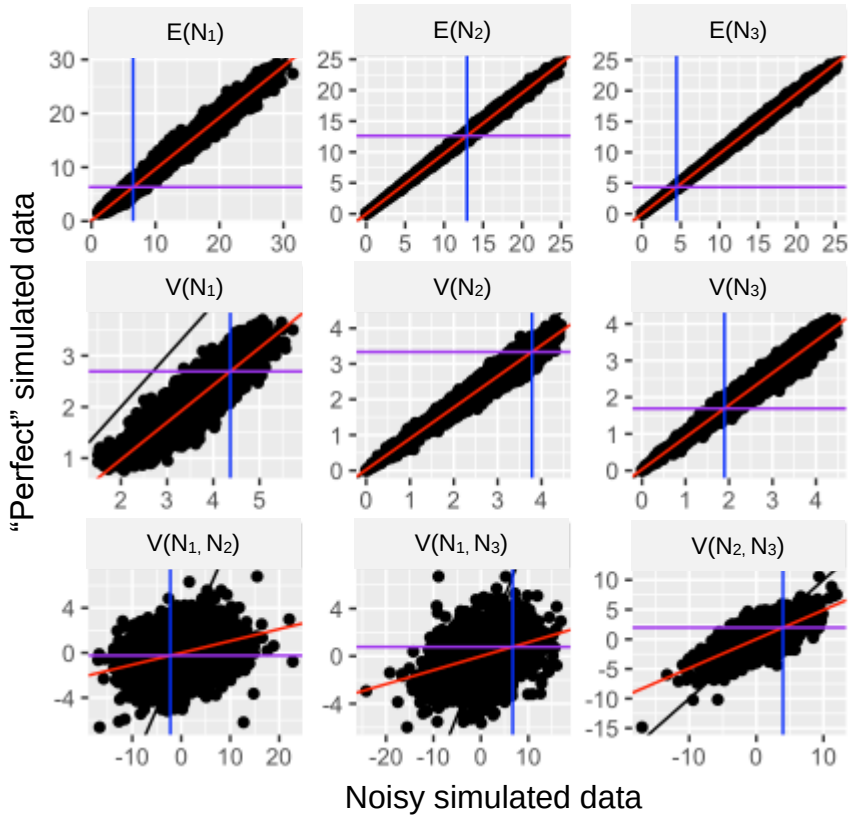


Figure 2.6: Linear regression models to quantify the relationship between noisy and "perfect" simulated data

These are plots generated by simulating "perfect" datasets and transforming them to "noisy" by applying the noise models by Equations 2.5.1 and 2.5.2, calculating the first- and second-order moments and plotting the "perfect" against the "noisy" moments. To capture the relationship between the noisy and "perfect" data for values close to the observed moments, a weighted linear regression model is fitted (red line), with points weighted proportionately to their proximity to the observed noisy moment in the experimental dataset. Using weighted linear regression models, it is possible to find the "perfect" moment (purple line) that corresponds to the observed "noisy" moment in the experimental dataset. For variances and covariances the values have been log-transformed. Negative covariance values were transformed into their absolute value prior to log-transformation; the sign of the covariance in the experimental data was then re-assigned to the "perfect" covariance value predicted from the weighted linear regression model.

2.6. Parameter Inference

Given a set of initial moments and a set of moments observed at a later timepoint, it is possible to find, a set of *predicted* moments that closely matches the observed moments at the later time point by minimizing the cost function that measures the distance between them; in this case, the KL-divergence (Section 2.4). The parameter set generating these well-matched predicted moments is selected as the best estimate of the unobserved processes which I aim to characterise and quantify.

To perform parameter inference, I used the dedicated SPEEDI.R package (Price *et al*., 2017), which offers a range of optimisation algorithms to minimise the cost function. In the absence of prior knowledge about the behaviour of the function to be optimised, I used the gradient-free Nelder and Mead optimisation routine (*optim* function in the *stats* package R) which is appropriate for non-smooth functions where derivative values are not available (Nelder and Mead, 1965). To abrogate the influence of the algorithm search starting point, I ran the optimisation routine 200 times on the same set of initial and endpoint moments with different starting points, and chose the parameter set with the lowest divergence value as the best.

2.7. Parametric bootstrap to assess model goodness-of-fit to the data

Goodness-of-fit is a measure of how well a given model captures the experimental data. Here, I used a parametric bootstrap approach to assess model goodness-of-fit. Using minimum-divergence moments-based inference, I identified the set of parameters that correspond to first- and second- order moments which yield the smallest divergence from the observed moments (detailed in 2.5). I refer to this set of moments as predicted moments and the corresponding parameter set as predicted parameters.

To evaluate the relative goodness-of-fit for the models, I used three metrics. The first metric was the pairwise comparison of the KL divergences; a lower divergence indicates a smaller discrepancy between the model-predicted moments and the experimentally observed moments. This metric is used to compare model performance overall, but is not informative about the model's ability to capture each individual moment. For example, it is possible for two models to have similar KL divergence values, which may, at first sight, suggest that the

two models perform similarly. However, upon examining the discrepancy between each individual experimental and model-predicted moment, it is possible to observe that one model captures a few individual moments very well, while completely misestimating the others, while the second model has a more homogeneous behaviour across all moments. In this case, I take the latter to be a better-performing model, as it captures the summary statistics adequately.

To assess the performance of the model at the level of the individual moments, I used two other metrics. For each model, I plotted the 95% confidence interval for each of the nine bootstrapped moments and overlaid the experimentally observed moments, which should fall within that confidence interval for a model with good fit, as explained in section 2.6. This metric assesses whether the model adequately captures all first- and second-order moments, which I have used as my summary statistics of choice. Figure 2.7 shows examples of models with poor (A), adequate (B) and good (C) fit.

The third metric was the pairwise comparison of the similarity between each synthetic dataset and its bootstrapped samples for the six models considered. Given the best parameter estimates after fitting each model to the experimental data, I used the parametric bootstrap approach to simulate 500 bootstrapped samples for each. I obtained the mean value and standard deviation of each moment for the 500 bootstrapped samples and calculated the number of standard deviations between each experimentally observed moment and the mean of each bootstrapped sample moment. This metric provides quantitative information about how close the experimentally observed moments are to the mean behaviour of the model considered.

For models with similar performance across the three metrics, I chose the one with the smallest number of parameters in line with the principle of parsimony.

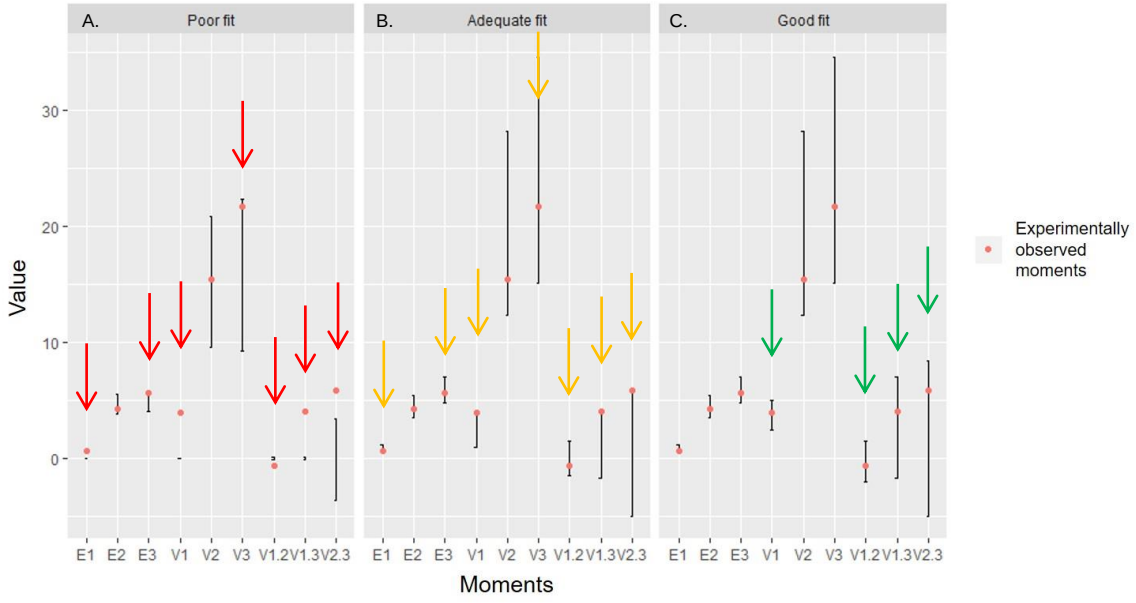


Figure 2.7: Bar plots of 95% confidence intervals for simulated moments from models with poor, adequate and good fit

The experimentally observed moments are represented by red dots, superimposed on bar plots of 95% confidence intervals for moments calculated from simulated datasets (N=200) using parametric bootstrap when three models are considered (A-C). Panel (A) represents a model with a poor fit to the data, as the 95% confidence intervals for the simulated moments differ significantly from the experimentally observed moments (red arrows). Panel (B) represents a model with adequate fit, with all experimentally observed moments captured only by very few simulations of the model (orange arrows), while Panel (C) represents a model with a good fit, with observed moments approximating the median values of the simulated moments.

3.

S. Typhimurium dynamics in the naïve host

Summary Points:

- In this chapter, I formulate and apply a series of nested compartmental models to two ITS-based datasets from intravenous *S.* Typhimurium infections in the naïve murine host with the aim of quantitatively characterising the unobserved demographic processes that shape bacterial dynamics in the naïve hots.
- In the naïve murine host, inter-tissue mixing is predicted to occur during the first 6 hours post-inoculation, which is earlier than previously thought.
- With parameter estimates for inter-tissue bacterial migration and intra-organ death and replication, the systemic phase of salmonellosis can be broken down to 3 distinct phases starting from inoculation: a) 6 hours of niche establishment in the organs and intra-organ killing, followed by b) 18 hours of exponential growth, and c) 48 hours of stable, marginal intra-organ growth.
- It is possible to identify the changes in demographic processes that underpin the transitions between the phases.
- In the early phase of infection, bacteria migrate asymmetrically to the liver and spleen, with higher migration to the former, where they are met with high bactericidal activity.
- In the second phase of infection, cidal activity in the liver and blood-to-organ migration decrease.
- In the third phase of infection, intra-organ bacterial dynamics remain stable, with bacterial replication higher than killing in both the liver and the spleen.

3.0. Abstract

In the 21st century, bacterial infections still constitute a major cause of morbidity and mortality disproportionately afflicting children and immunocompromised individuals in low-resource settings. Of bacterial diseases, systemic salmonellosis remains a leading cause of community-acquired bloodstream infections in sub-Saharan Africa. *Salmonella* Typhimurium infections of the murine host have long served as an experimental model for human systemic salmonellosis to investigate virulence determinants, host immune responses and within-host bacterial population dynamics amongst others. Despite the continuous improvement of experimental techniques, *in vivo* observation of bacterial replication, killing and inter-tissue migration remains imperfect. Combining experimental data with mathematical modelling can address this challenge by enabling inferencing about experimentally unobserved processes. Here, a mechanistic, stochastic mathematical model is fitted to data obtained from experiments with isogenic tagged *Salmonella* strains to reveal that, in the early stages of infection, bacteria freely migrate out of the liver and spleen and are redistributed via the bloodstream. Estimation of the replication, killing and migration rates shows differential *Salmonella* uptake in the organs until 24 hours post-infection, before the rates stabilise and converge to similar levels.

3.1. Introduction

Infections of genetically susceptible mice with *S.* Typhimurium constitute one of most extensively studied host-pathogen system since the 1980's, leading to insights in the fields of *Salmonella* pathogenesis, immunity and vaccinology, while also serving as an accelerator for the development of new methods in microbiology (reviewed in Tsolis *et al*., 2011). The wealth of information about the genes and systemic mediators that affect the course of *Salmonella* infections in mice is a valuable resource that has allowed incremental discoveries to be made. Examples include the discovery that *aroA* Salmonella mutants are attenuated and immunogenic in mice (Hoiseth and Stocker, 1981), which was later used by Rossi *et al.* (Rossi *et al*., 2017) to compare the effect of antibiotics in rapid-growing and slow-growing (*aroA* mutant) strains. Similarly, the finding that *Salmonella* infection control requires both antibodies and $CD4^+$ T-cells inspired the comparison between the mechanisms of action of live-attenuated and killed whole-cell vaccines (Coward *et al*., 2014).

In contrast to the rich literature around the host immunological and bacterial virulence factors and their effect on *Salmonella* colonisation, pathogenesis and response to vaccination, little is known about the events that occur *in vivo* in terms of when, where and how quickly bacteria replicate, die and migrate between tissues in the naïve murine host (Sheppard *et al*., 2003). Studies using confocal microscopy (Richter-Dahlfors, Buchan and Finlay, 1997; Sheppard *et al*., 2003) and ITS (Grant *et al*., 2006; Kaiser *et al*., 2013) have provided a series of static snapshots regarding the intracellular distribution of *Salmonellae*, the host cell types *Salmonellae* associate with, the tissues/organs colonised and the bacterial burden in each.

However, insights into the mechanisms underpinning these static observations are lacking. One way of completing the picture is to integrate mechanistic mathematical models with experimental snapshot data in order to make inferences about the unobserved biological processes of bacterial replication, death and migration that collectively shape the fluctuations in the size of the bacterial subpopulations in different tissues. A study by Brown *et al*. (Brown *et al*., 2006) using mechanistic models to re-analyse *in vivo* data on intracellular *Salmonella* counts in liver macrophages (Sheppard *et al*., 2003) is a characteristic success story of using this approach, by developing a series of models to address the question: which modes of intracellular replication and host cell lysis are compatible with the observed numbers of bacteria in liver macrophages?

Similarly, it is possible to use this integrated mathematical and experimental approach to address questions about the dynamics of systemic *S.* Typhimurium infections in the naïve host at the level of the bacterial population. For example, the role of both humoural and cell-mediated immunity in the early control of Salmonella proliferation in the blood of the murine host remains contested (Mittrücker and Kauffman, 2000), especially for systemic infections acquired intravenously, in which bacteria do not go through the immunologically active gut microenvironment, thus having a lower probability of encountering circulating phagocytes and becoming internalised. Previous modelling attempts using ITS-based data from systemic *S.* Typhimurium infections in the naïve murine host have *a priori* ignored the possibility bacterial killing in the blood (Dybowski *et al*., 2015; Coward *et al*., 2014; Grant *et al*., 2008). Another uncertainty concerns the ways in which *Salmonellae* colonise the systemic tissues.

Previous studies have posited that bacteria colonise the liver and spleen, forming local foci which remain segregated until later in the infection process (Grant *et al*., 2008; Dybowski *et al*., 2015); however, these studies only used simplified models with as few parameters as possible for feasibility purposes, and did not comparatively assess the goodness-of-fit of rival more complex models.

Other unknowns include the ways in which different tissues control infection locally, e.g. whether they rely on bactericidal or bacteriostatic mechanisms and the extent of disparities between the different tissues. Finally, there is a knowledge gap around the later phases of systemic salmonellosis, where bacterial subpopulations grow exponentially in both the liver and the spleen: is this phenomenon the result of a higher average bacterial replication rate, or lower cidal activity in the organs?

Deciphering these questions is important not only to gain a more complete understanding around the pathogenesis and host-bacterial interactions in systemic salmonellosis, but also to provide a quantitative reference which will enable the characterisation of the therapeutic or prophylactic effects of interventions such as antibiotic treatment or immunisation. In this chapter, I use ITS-based datasets from two experiments combined with data-driven mechanistic models to address the questions outlined earlier with regards to the within-host dynamics of *S.* Typhimurium in the naïve host. First, I formulate a series of mechanistic models that reflect rival biological hypotheses regarding the early inter-organ bacterial mixing, and killing and replication of *Salmonellae*, as well as the host-bacteria interactions in the later phases of systemic salmonellosis. Then, I evaluate their performance by assessing

their goodness-of-fit to the ITS-based data, and reject those that do not capture the patterns in the experimental data. Finally, following model selection, I perform parameter estimation and provide confidence intervals for the inferred parameters.

3.2. Methods

3.2.1. Experimental Data Background

Coward *et al.* (Coward *et al.*, 2014) collected two datasets on bacterial counts of isogenic-tagged *S.* Typhimurium SL3261 strains in C57BL/6 mice. Mice were infected with a ~240 CFU *S.* Typhimurium equiproportionate inoculum comprising 8 ITS at ~30 CFU per strain. At 0.5, 6, 24, 48 and 78 hours post-inoculation, 9-10 mice were killed, their livers and spleens harvested, blood was sampled from their tail vein, bacteria isolated and segregated by strain using qPCR. By studying the changes in ITS composition in the tissues with time and comparing strain distributions between tissues it is possible to make some preliminary inferences about the within-host dynamics of *S.* Typhimurium in the naïve murine host. Figure 3.1 illustrates the number of bacteria per ITS in each of the biological replicates (9 or 10 mice x 8 ITS/mouse = 72 or 80 total biological replicates) for each sampling time point.

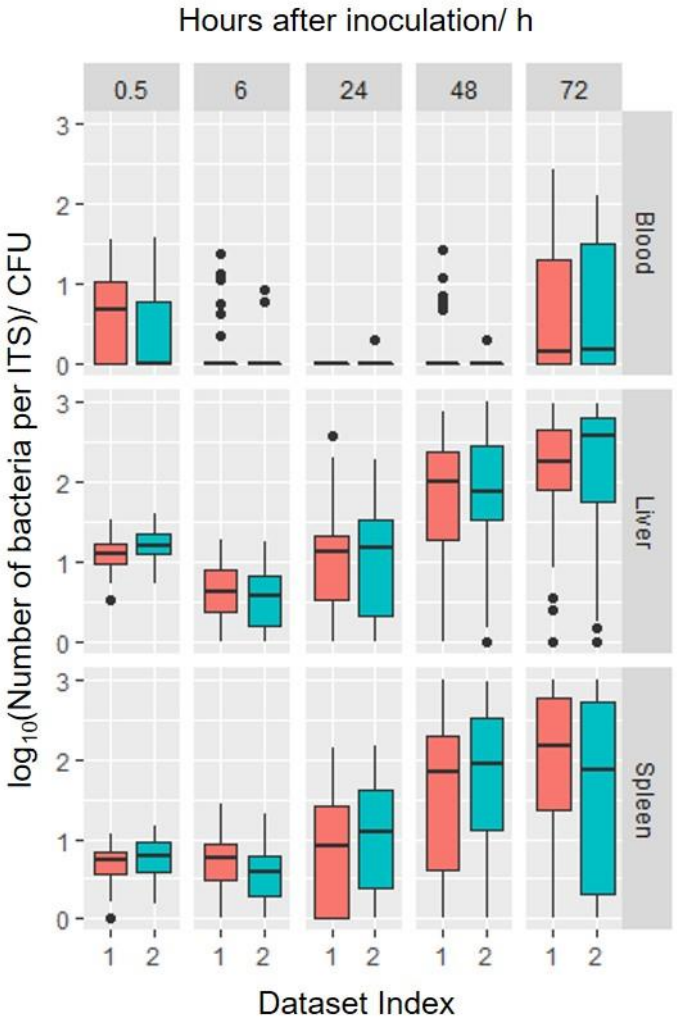


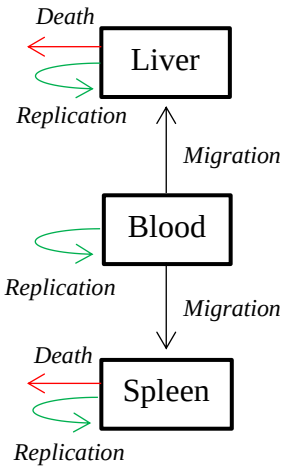
Figure 3.1: ITS composition of bacterial loads in different organs at various timepoints post-infection.

This is a boxplot representation of the two datasets collected from infections of naïve mice with an inoculum comprising 8 ITS at ~30 CFU per ITS. Each boxplot represents the distribution of copies per ITS in the tissues sampled, the blood, liver and spleen for five sampling time points over the span of 72 hours. Comparing datasets 1 and 2, there is concordance in the observed bacterial loads. In the blood, bacterial loads diminish between 6 and 48 hours before increasing again by 72 hours. In the liver, bacteria decline between 0.5 and 6 hours but thereafter increase steadily. In the spleen, the bacterial load remains static for the first 6 hours and increases from 6 to 72 hours, similarly to the liver.

3.2.2. Mathematical models and corresponding biological hypotheses for early S. Typhimurium dynamics

In this section, I present a series of nested mechanistic mathematical models considered and their correspondence to biological hypotheses suggested by previous experimental studies or patterns observed in the datasets (Figure 3.1).

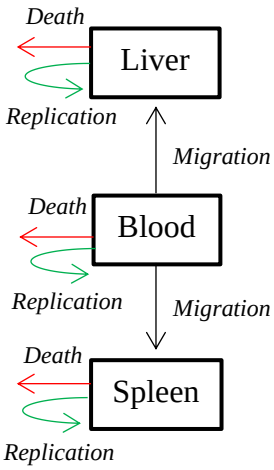
- **M1: No early inter-organ bacterial mixing with no bacterial death or replication in the blood**



An ITS-based study by Grant *et al.* (2008) recovered bacteria from the blood, liver and spleen, and negligible numbers in other organs (kidneys, lymph nodes, heart and lungs. An original mathematical model for systemic salmonellosis was formulated including 3 compartments (blood, liver and spleen). Bacteria migrated from the blood to the liver and spleen, with bacterial replication and death in the same organs, as shown in Figure 3.2. The same model was later used by Coward *et al.*, 2014. This simplified model does not include bacterial replication and killing in the blood.

Figure 3.2: Conceptual structure of early compartmental model M1 for systemic salmonellosis in the murine host

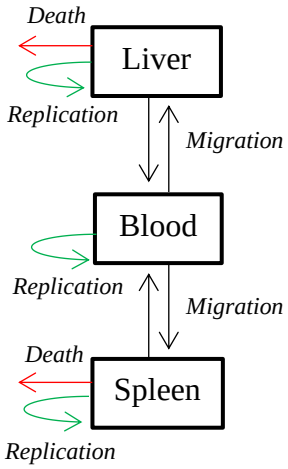
- **M2: No early inter-organ bacterial mixing with bactericidal activity in the blood**



For an infection acquired intravenously, it is unclear how soon *Salmonella* enters host cells and whether it circulates as a free bacterium, attached to or phagocytosed by circulating host cells. This model includes bactericidal activity in the blood in the early stages of systemic infection, reflecting either humoural or cellular responses (Figure 3.3). Possible improvement in the model fit between M1 and M2 would provide support for an early cidal activity in the blood.

Figure 3.3: Conceptual structure of early compartmental model M2 for systemic salmonellosis in the murine host

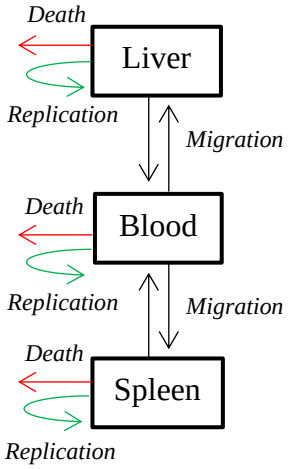
- **M3: Early inter-organ bacterial mixing without bactericidal activity in the blood**



The presence of low-level bacteraemia at 6h in the datasets by Coward *et al.* (2014) (Figure 3.1) led me to the hypothesis that there is free inter-organ mixing between the liver and spleen, as bacteria may escape in the bloodstream earlier than previously thought (Figure 3.4), and in contrast to suggestions made by Grant *et al.*, 2008.

Figure 3.4: Conceptual structure of early compartmental model M3 for systemic salmonellosis in the murine host

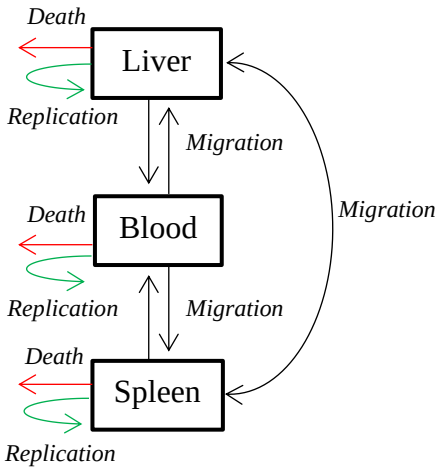
- **M4: Early inter-organ bacterial mixing with bacterial death in the blood**



In M4, I consider a biological scenario with both free inter-organ bacterial mixing, and early bactericidal activity in the blood (Figure 3.5).

Figure 3.5: Conceptual structure of early compartmental model M4 for systemic salmonellosis in the murine host

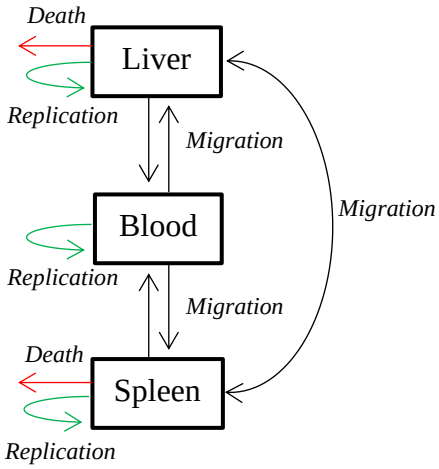
- **M5: Early inter-organ bacterial mixing in systemic circulation and the liver-spleen axis with bacterial death in the blood**



M5 represents an extension of M4 taking into account more than one possible routes for bacterial migration between the liver and spleen (Figure 3.6). In the datasets used, blood was fractionally sampled from the tail vein, which represents a distal location and may not be representative of fast and direct inter-organ haematogenous spread via shorter routes, like the liver-spleen axis which forms part of the portal circulation.

Figure 3.6: Conceptual structure of early compartmental model M5 for systemic salmonellosis in the murine host

- **M6: Early inter-organ bacterial mixing in systemic circulation and the liver-spleen axis with no bacterial death in the blood**



In M6, I consider a biological scenario with free inter-organ bacterial mixing both through the systemic and portal circulation, but no bactericidal activity in the blood (Figure 3.7).

Figure 3.7: Conceptual structure of early compartmental model M6 for systemic salmonellosis in the murine host

3.2.3. Model Selection

Each of the six candidate models for the early *S.* Typhimurium dynamics (outlined in 3.2.2) was fitted to the experimental data using the first two sampling time points at 0.5 and 6 hours after inoculation.

To evaluate the relative goodness-of-fit for the models, I used the three metrics described in section 2.7. In summary, the first metric is the pairwise comparison of the KL divergences and is an index of the overall model performance, without providing information about how well the model captures the individual experimental moments. For the second metric, I plotted the 95% confidence interval for each of the nine bootstrapped moments and overlaid the experimentally observed moments. This metric assesses whether the model adequately captures all moments; a model with a good fit should adequately capture all individual moments. Finally, for the third metric I calculate the standard deviation and mean for each of the nine moments in the bootstrapped samples, and calculate how many standard deviations each experimentally observed moment differs from the mean value of each bootstrapped moment. The smaller the number of standard deviations, the closer the experimentally observed moments are to the mean behaviour of the model.

For models with similar performance across the three metrics, I chose the one with the smallest number of parameters in line with the principle of parsimony.

3.2.4. Parameter Inference

Parameter inference for each model was performed by minimisation of the KL divergence between the set of experimentally observed moments and predicted moments given a parameter set θ, as described in section 2.5.

3.2.5. Quantification of uncertainty around parameter estimates

To quantify the uncertainty around the parameter estimates, I used a parametric bootstrap approach (outlined in 2.5 and 2.6) to generate 500 bootstrapped datasets from the best-performing model. For each bootstrapped dataset, I calculated the moments and performed inference by fitting the model and obtaining a set of best parameter estimates. For each parameter, I plotted the 95% confidence interval.

3.3. Results

3.3.1. Model Selection for the early phase dynamics (0 to 6 hours)

In section 3.2.2, I formulated a series of mechanistic models inspired by insights from previous biological studies and/or patterns observed in experimental data. Here, I compare the performance of the six models (M1-M6) based on the three criteria described in section 3.2.3. Table 3.1 shows the KL divergence values for models M1 to M6, with a large drop seen for M3, which introduces bacterial migration from the liver (m21) and spleen (m31) to the blood. Early cidal activity in the blood (k1) does not substantially improve the congruence between the model prediction and the experimental observations (M2). Models more complex than M3 do not improve the model fit, as shown by the near-identical KL divergence values (M4-M6). A full description of the model parameters is provided in Table 3.2.

Model name	Parameters in the model	KL divergence (dataset 1)	KL divergence (dataset 2)
M1	m12, m13, k2, k3, r1, r2, r3	4.30	5.91
M2	m12, m13, k1, k2, k3, r1, r2, r3	4.21	5.56
M3	m12, m13, **m21, m31**, k2, k3, r1, r2, r3	0.37	0.27
M4	m12, m13, m21, m31, k1, k2, k3, r1, r2, r3	0.34	0.26
M5	m12, m13, m21, m31, m23, m32, k1, k2, k3, r1, r2, r3	0.35	0.27
M6	m12, m13, m21, m31, m23, m32, k2, k3, r1, r2, r3	0.35	0.28

Table 3.1: A comparative table of the KL divergence values for models M1 – M6

	Parameter name	Description
Inter-organ migration	m12	Bacterial migration from the blood to the liver
	m13	Bacterial migration from the blood to the spleen
	m21	Bacterial migration from the liver to the blood
	m31	Bacterial migration from the spleen to the blood
	m23	Bacterial migration from the liver to the spleen
	m32	Bacterial migration from the spleen to the liver
Intra-organ dynamics	k1	Bacterial death in the blood
	k2	Bacterial death in the liver
	k3	Bacterial death in the spleen
	r1	Bacterial replication in the blood
	r2	Bacterial replication in the liver
	r3	Bacterial replication in the spleen

Table 3.2: Description of the model parameters

In terms of the model predictions for individual moment values, models M1 and M2 both underestimate the variance in the blood when applied to the two experimental datasets, while models with organ-to-blood bacterial migration (m21, m31) address this mismatch (Figure 3.8). In Figure 3.8, I plot the 95% confidence intervals for the moment values, as estimated from 500 bootstrap samples generated from each of the models M1-M6 originally fitted to two experimental datasets (Panels A and B). Consistent with the pattern in the KL divergence values in Table 3.1, Figure 3.8 shows that M3 adequately captures all experimentally observed moments, while models with higher complexity (M4-M6) also converge to the same pattern.

To further quantify the alignment between the model-generated predictions and the experimental observations for each moment individually, I assess the proximity of each experimental model to the mean model prediction for each moment. To quantify this distance, I use first calculate the mean value and standard deviation for each moment using the bootstrapped samples across the six models, and report the difference between each experimental and model-predicted moment as the number of standard deviations away from the mean. A decrease in the number of standard deviations reflects a model whose mean behaviour is closer to the experimental observations. In Figure 3.9, I plot the $\log_{10}$ of standard deviations away from the mean for each moment across the six models and two experimental datasets. The results are consistent with the patterns noted earlier (Table 3.1 and Figure 3.8), with M3 providing a significant improvement, but subsequent models of higher complexity providing only marginal further improvements.

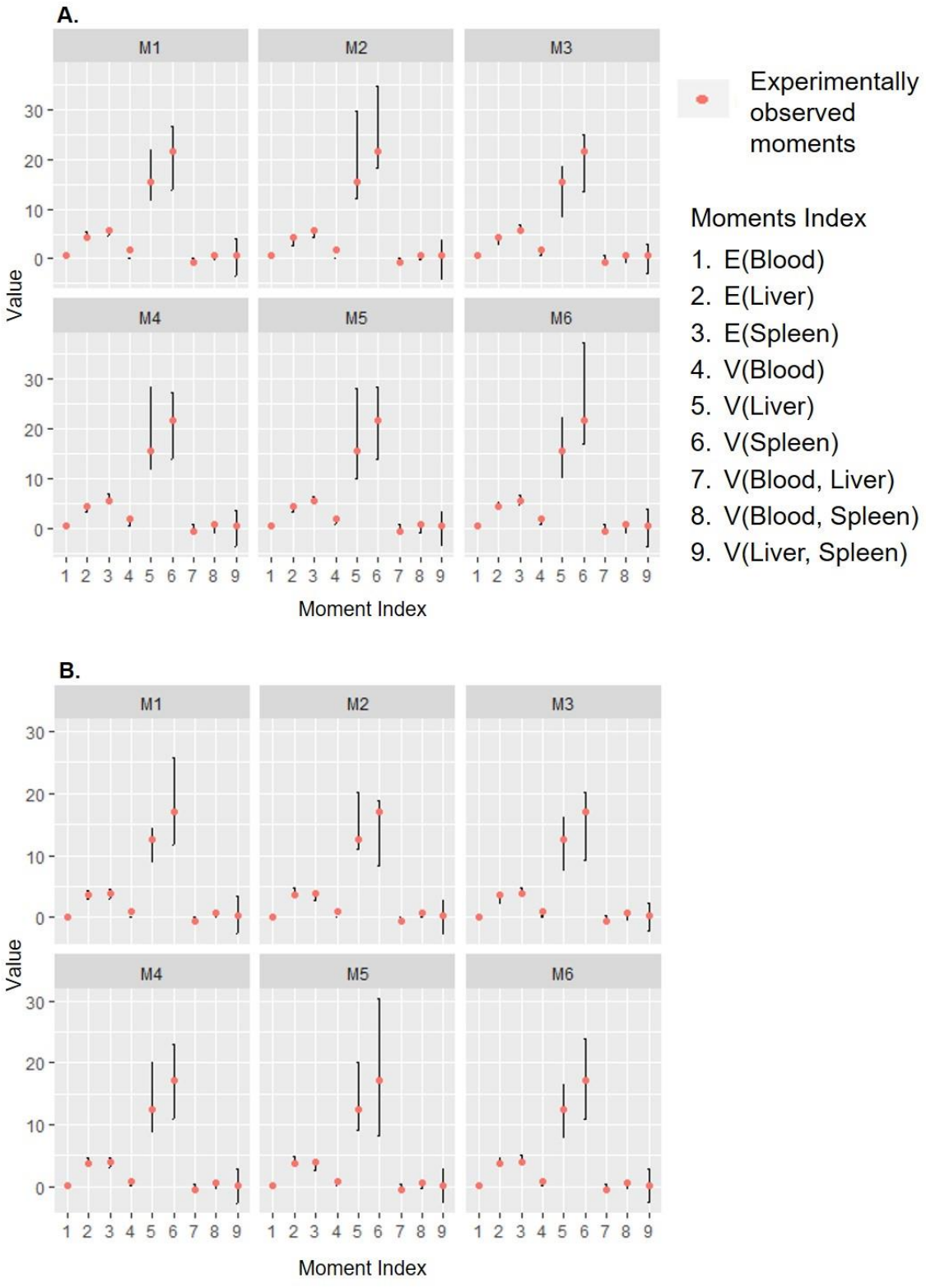


Figure 3.8: Bar plots of the 95% confidence intervals for the moment values, as estimated by parametric bootstrap ($n = 500$) using models M1-M6

Panels A and B correspond to two experimental datasets. For each moment, the 95% confidence interval obtained by parametric boostrapping is shown as a bar plot. The red dots represent the experimentally observed moments. In both (A) and (B), M1 and M2 consistently underestimate V(Blood), while M3 improves the fit. M4-M6 follow the same pattern as M3.

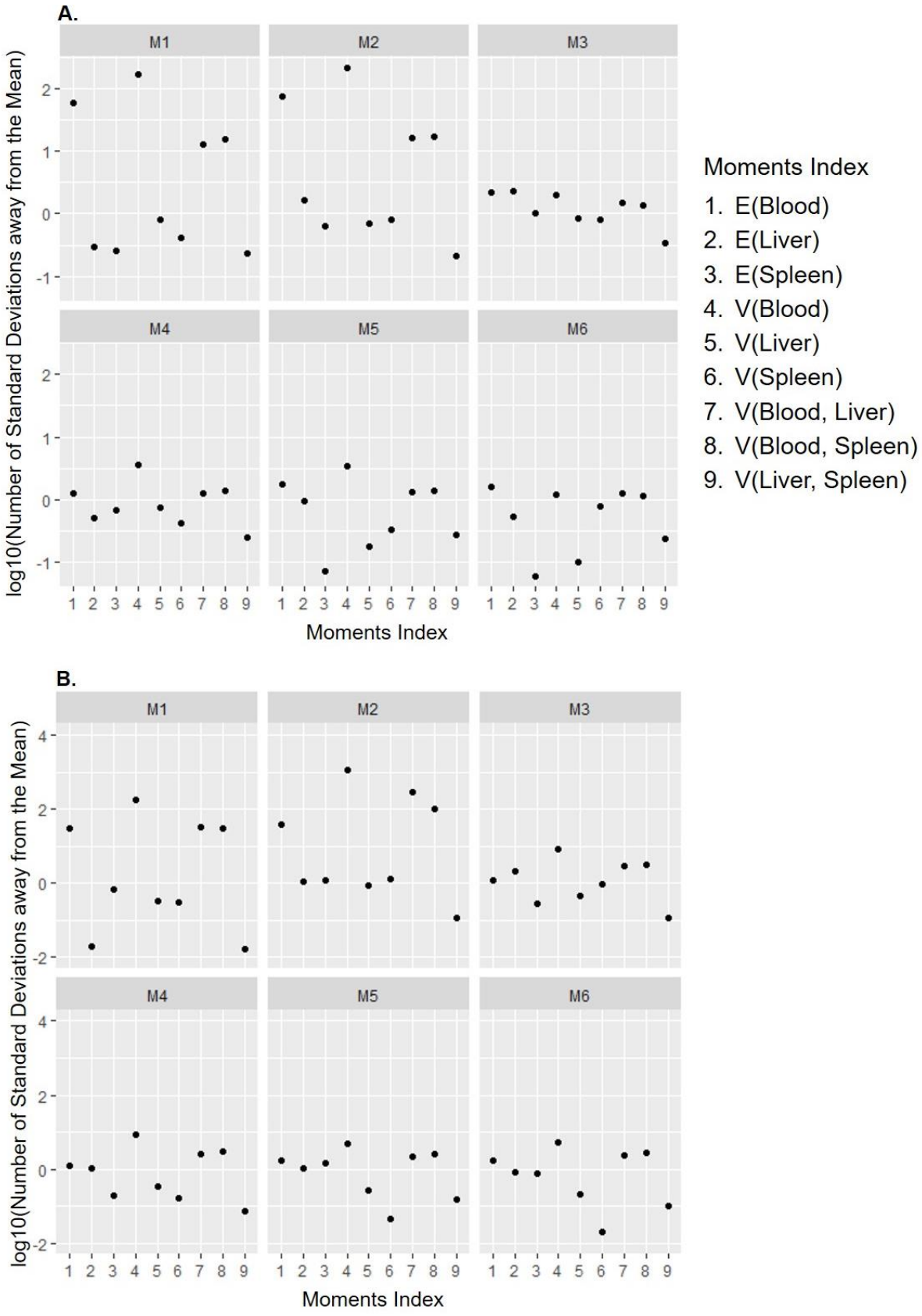


Figure 3.9: The difference between the experimentally observed and model-predicted moment values expressed as the $\log_{10}$ of the number of standard deviations away from the mean ($n = 500$)

3.3.2. Quantification of within-host dynamics in the naive host

To discriminate the processes of within-host bacterial replication, killing and migration, I fit M3 to the two experimental datasets by Coward *et al.* (2014) for a course of 72 hours in the blood, liver and spleen of naive mice. I inferred the rates at which these processes occur by identifying the parameters, whose corresponding first- and second-order moments have the smallest distance from the moments characterising the distribution of observed bacterial counts per ITS. I obtained confidence intervals by using a parametric bootstrap approach, as outlined in section 2.5. In the naive host, independent model fitting for the two experimental datasets revealed very similar patterns of infection dynamics throughout the first 3 days post-inoculation, both in terms of temporal evolution and in terms of the relative contribution of bacterial killing, migration and replication to the observed bacterial numbers. This congruence provides a degree of external validation to the inference conclusions.

- **0-6 hours**

In the early stages of infection there is a marked preferential migration of bacteria from the blood to the liver (high predicted m12). At this stage, the liver is seeded at a higher rate than the spleen (m12 > m13). Bacteria escape back to the blood at a lower rate (non-zero, but small m21, m31 values). In terms of intra-organ activity, the bactericidal activity of the liver is high and exceeds the rate at which bacterial replicate (k2>r2), resulting in net bacterial death locally. In contrast, the bactericidal activity is much lower in the spleen (k3<k2) allowing low-level bacterial proliferation (k3 < r3) in that organ (Figure 3.10).

- **6 to 24 hours**

In the second phase of the infection, the migration of bacteria from the blood to both the liver and spleen slows down (low rates of m12 and m13). No bacteria are detected in the blood at 24 hours, as bacterial subpopulations remain segregated in the liver and spleen (m21, m31 close to zero). Hepatic bactericidal activity drops, while that of the spleen remains stable. Bacterial replication accelerates in the liver, matching the sustained pattern observed in the spleen post-inoculation, resulting in replication rates marginally higher than killing rates and overall net intra-organ growth of the bacterial in both organs (Figure 3.11).

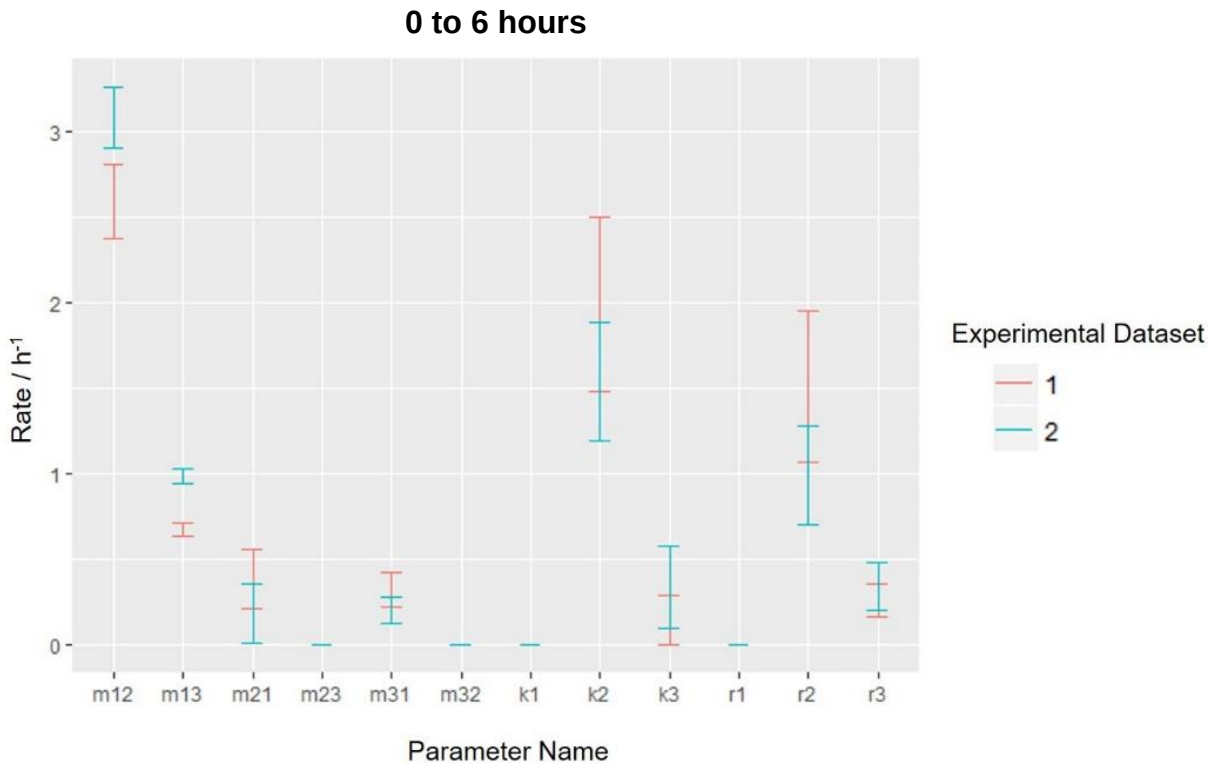


Figure 3.10: Bar plots of the 95% confidence intervals for the parameter values during the acute phase of infection (0 to 6 hours), as estimated by parametric bootstrap ($n = 500$) using model M3

The red bar plots represent the 95% confidence intervals for the parameter estimates obtained by fitting M3 to the first experimental dataset, while the blue bar plots represent the parameter estimates obtained by model fitting to the second experimental dataset. The estimates obtained from the two datasets show congruent patterns in the bacterial dynamics, as shown by the overlapping bar plots. Overall, the majority of bacteria migrate to the liver, where they are subject to intra-organ cidal activity, with a small population escaping back to the blood. The minority of bacteria that migrate to the spleen grow at a low rate (mean net growth in the spleen = 0.156 h^{-1})

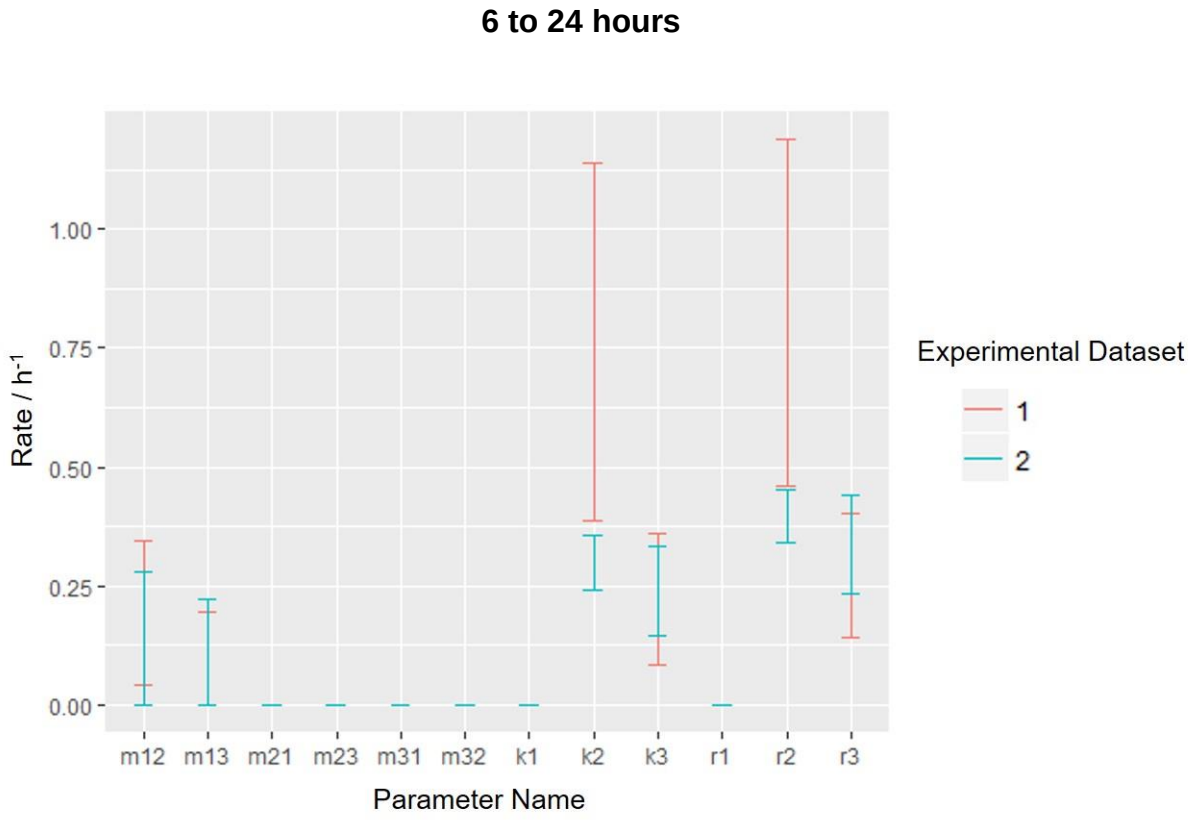


Figure 3.11: Bar plots of the 95% confidence intervals for the parameter values during the second phase of infection (6 to 24 hours), as estimated by parametric bootstrap ($n = 500$) using model M3

The red bar plots represent the 95% confidence intervals for the parameter estimates obtained by fitting M3 to the first experimental dataset, while the blue bar plots represent the parameter estimates obtained by model fitting to the second experimental dataset. The rate of bacterial killing (k2) in the liver drops, creating an environment permissive to bacterial proliferation. The large confidence for the replication and killing rates (k2, r2) for the second experimental dataset reflect highly imprecise estimates attributable to high data variance. Dynamics in the spleen remain similar to earlier (0-6 hours), with continuing intra-organ net positive growth.

- **24 to 72 hours**

In the third phase of the infection (24-72 hours), there is sustained bacteraemia with increasing bidirectional bacterial migration from and to the blood (m12, m13, m21, m31 rates are higher compared to the second phase in Figure 3.11). Bacterial replication (r2, r3) and death (k2, k3) remain at similar levels as before in the liver and spleen, while bacteria replicate in the blood (non-zero r1), as seen in Figure 3.12.

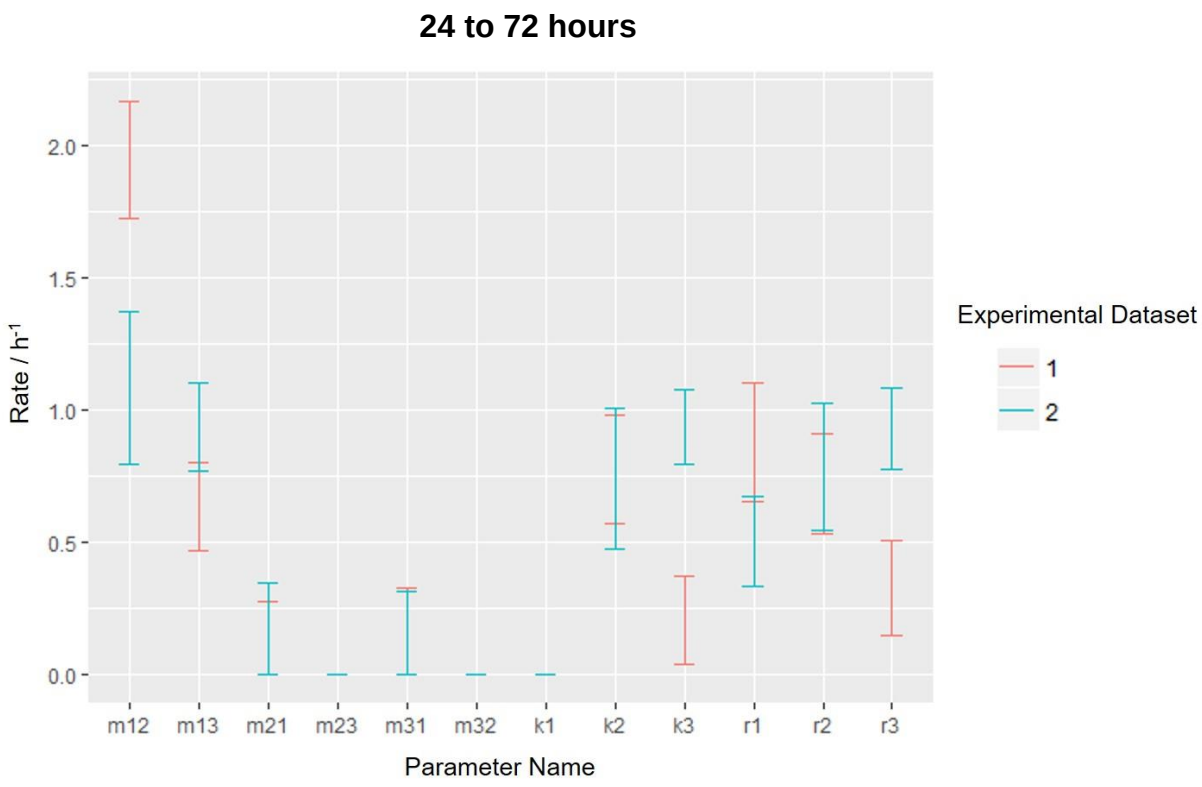


Figure 3.12: Bar plots of the 95% confidence intervals for the parameter values during the late phase of infection (24 to 72 hours), as estimated by parametric bootstrap ($n = 500$) using model M3 with bacterial replication in the blood (r1)

3.4. Discussion

In this chapter, I applied a newly described inference framework to published experimental datasets by Coward *et al.*, 2014 to quantitatively characterise the within-host dynamics of intravenously acquired *S.* Typhimurium infections in the naive murine host. My modelling work suggests that the early infection phase is characterised by rapid migration of bacteria from the blood to the liver and spleen, with early inter-organ bacterial mixing. The intra-organ microenvironments differ between the liver and spleen; while bacteria migrating to the liver are subject to high levels of cidal activity, bacteria lodging in the spleen are allowed to proliferate. In the second phase of infection (6 to 24 hours), the hepatic bactericidal activity drops, creating a more permissive environment for bacterial growth, which continues in the later phase of infection (24-72 hours) in both the liver and spleen.

The modelling work presented here is, to the best of my knowledge, the first attempt in the field of population dynamics microbiology that successfully integrates mathematical models with experimental data for the entire course of an experimental infection process (72 hours). Although there is a wealth of experimental studies investigating the effect of virulence and/or immunological factors on the early colonisation, dissemination and proliferation process that collectively shape the infection dynamics in the naïve host, insights into the unobserved temporal dynamics resulting from host-bacterial interactions have been hitherto lacking, due to three factors: 1) technological limitations not allowing the direct observation and tracking of replication, killing and migration events at the level of the entire bacterial population, 2) the previous absence of computationally efficient mathematical tools to perform inference, and 3) a focus on systemic infections acquired via haematogenous spread of orally-acquired *S.* Typhimurium infections, as opposed to blood-acquired infections.

Earlier studies have partially shed light on the early-phase dynamics, identifying the spleen and liver as major bacterial reservoirs (Grant *et al*., 2008), where bacteria reside intracellularly, replicate, lyse their host cell and generate new infectious foci by infecting adjacent cells (Sheppard *et al*., 2003; Brown *et al*., 2006). My inference suggests that the liver plays an important role in the *Salmonella* dynamics during the acute phase of infection, with the rate of blood-to-liver bacterial migration estimated to be three times higher than the rate of blood-to-spleen migration. For the early stages of intravenously acquired infections, it is unknown whether bacteria are free or reside intracellularly. Although it has been

suggested that immediately after gaining access to the blood, *Salmonella* swiftly invades macrophages of the reticuloendothelial system (Mastroeni, 2002), it is unclear whether this applies only to bacteria acquired through the orogastric route prior to haematogenous spread. Furthermore, it remains unknown whether bacterial movement at this stage is chemotaxis- or blood-flow dependent. The liver and spleen harbour different proportions of the host cell types that *Salmonellae* typically reside in (e.g. macrophages and dendritic cells) (Johansson and Wick, 2004), rendering chemotaxis-dependent bacterial migration a possible explanation for the asymmetric bacterial migration estimated during the first few hours post-inoculation. A cellular makeup exerting stronger chemotaxis in the liver may underpin the high blood-to-liver migration rates in the early stages of infection. An alternative explanation would be the differential blood perfusion of the two organs. Gjedde and Gjeode (Gjedde and Gjeode, 1980) report that in BALB/c mice, the liver is one of the most profusely perfused organs with allocated heart output of 945mL/min, as opposed to the poorly perfused spleen with 46mL/min.

In terms of model selection for the early phase of infection (0-6 hours), I showed that previously used models omitting the organ-to-blood bacterial migration (m12, m13) provide a poor fit to the experimental data (Grant *et al*., 2008; Coward *et al*., 2014); addition of a parameter for early bacterial killing (k1) in the blood does not improve the model fit either. On the basis of early modelling attempts in a maximum likelihood inference framework with a simplified model, Grant *et al*. (2008) argued that once bacteria migrate to the liver and spleen, they remain trapped in their respective microenvironments until much later in the infection process. My analysis in 3.1 suggests the opposite, i.e. that early inter-organ mixing is necessary to reproduce the distribution of copies per ITS in the experimental datasets. At this stage and in the presence of a large population of uninfected monocytes and macrophages in the liver and spleen, it is likely that *Salmonellae* become internalised by host cells. Once internalised, *Salmonellae* have been reported to alter the surface electrical properties of host phagocytes in the gut (Sun *et al.*, 2019; Worley *et al*., 2006), which in combination with the electrochemical properties of their microenvironment can lead to directional movement of *Salmonella*-infected macrophages. A similar mechanism may underpin the early inter-organ migration which occurs in tandem with intra-organ expansion of infection foci in the systemic phase of infection (Brown *et al.*, 2006).

Furthermore, my analysis in 3.3.1 shows that including a parameter for early bactericidal activity in the blood does not substantially improve the model fit. This result is consistent with the finding that upon infection with *Salmonella*, mouse serum deposits only a small amount of C3 complement onto the bacteria and, as a result, exerts no bactericidal activity (Siggins *et al*., 2011).

Finally, the results presented in 3.3 pinpoint and quantify the changes that signpost the transition between the three phases of infection in the 72-hour period. A drop in bactericidal activity in the liver, as well as relative quiescence in the blood-organ communication, both signal the transition from the first (0-6 hours) to the second (6-24 hours) phase of the infection. Between the second (6 to 24 hours) and third (24-72 hours), the intra-organ dynamics remain fairly stable, with bacteria migrating out of the organs into the blood. The inferred dynamics in the blood also present an interesting pattern; from a neutral microenvironment in the first phase (k1, r1 = 0), to a degree of cidal activity in the second, and bacterial replication in the third. These dynamics are likely to reflect the changing patterns in host cell-bacterial interactions, as the bacteria are likely to be extracellular in the first phase, and occupy different types of host cells between the second and third phases.

Although the modelling approach proposed by Coward *et al.* (2014), and extended here has enriched our insight and provided evidence in favour or against long-standing speculations, it is, nonetheless, subject to assumptions, whose acknowledgement can guide future experimental and modelling work. As each ITS is assumed to represent a single independent infection, the case of 10 mice each infected with 8 ITS represents $8 \times 10 = 80$ stochastic trajectories of a single infection, an interpretation underpinned by 3 assumptions outlined below:

1. that all 8 WITS are phenotypically identical and exhibit identical within-host dynamics,
2. that bacteria behave independently of each other,
3. that host-bacterial dynamics are uniform across all mice, with no inter-mouse heterogeneity,
4. that the rates at which demographic events take place are constant for the modelled time interval.

Of course, there is a wealth of studies showing that bacterial dynamics *in vivo* systems are complex and heterogeneous. For example, Helaine *et al.* (Helaine *et al.*, 2010) and Claudi *et al.* (Claudi *et al.*, 2014) argue that the composition of a bacterial population is not uniform and is characterised by diversity in both division rates and susceptibility to killing possibly due to bacteria residing in different microenvironments and acquiring of epigenetic marks during the infection process, while Balaban *et al.* (Balaban *et al.*, 2004) suggest that bacterial heterogeneity is present in the bacterial culture even prior to inoculation of the host.

Regarding heterogeneity in host-pathogen interactions, data from Coward *et al.* (Coward *et al.*, 2014), showed that at 6 hours post-infection, 2 out of 10 mice harbour 20 and 35 CFU respectively in their blood, while no bacteria were detected in the blood samples collected from the remaining 8 mice. This bimodality in bacterial numbers, which is not necessarily adequately captured in the respective first- and second- order moments when all 10 mice are considered as a group, may reflect heterogeneity in host responses to infection. While acknowledging the presence of these additional layers of complexity is important to critically appraise the inferences made when using mathematical models, it is equally important to remember that the complexity of models is limited by data resolution and availability, and that the law of parsimony should be obeyed once a model adequately captures the patterns in the experimental observations.

Overall, in this chapter I used a likelihood-free inference framework to quantify the within-host dynamics of intravenously-acquired *S.* Typhimurium in naive murine host over a period of 72 hours. My analysis rejects the earlier conjecture that upon inoculation bacteria remain trapped in the liver and spleen during the early stages of infection, and suggests that there is early mixing between the two populations with the liver acting as a net recipient. Furthermore, my inferences show that after 6 hours, bacterial replication and killing rates in the liver and spleen do not change significantly, while maintaining a marginally positive net growth rate, which allows bacteria to grow exponentially. The rates reported in this analysis represent population averages and do not take into account inter-host or inter-bacteria phenotypic heterogeneity. To assess the effect of inter-host heterogeneity, one would need to obtain time series data from different animal hosts; this could be achieved by visualising bacteria during the infection timeline instead of culling and sampling different animals at each time point. Hierarchical models could be used to capture inter-host heterogeneity too. Finally, to address the issue of bacterial heterogeneity, observations of demographic events

at the level of the single bacterium with technologies like confocal/fluorescence microscopy or flow cytometry of previously tagged bacteria are needed.

4.

S. Typhimurium dynamics in the immunised host

Summary Points:

- Following from Chapter 3, I apply a compartmental model to ITS-based datasets from groups of mice previously immunised with either a live-attenuated (LV) or a whole-cell killed vaccine (KV), and challenged intravenously with *S.* Typhimurium.
- The LV group rapidly inactivates 1/3 of the inoculum within 30 minutes post-inoculation, while such effect is not observed for the KV and naïve groups.
- In the first 6 hours post-inoculation, both the KV and LV groups exert bacteriostatic effects in the liver with enhanced bactericidal activity in the spleen, compared to the naïve group.
- In the second phase of infection (6-24 hours), the bacteriostatic effect persists in the livers of both KV and LV mice. In the spleen, the bactericidal activity of the naïve group increases.
- In the third phase of infection (24-72 hours), the LV group controls intra-organ proliferation for the longest time, while in the KV group positive net bacterial growth is observed at 48 hours.

4.0. Abstract

With the rise of antibiotic resistance and tolerance, relapsing bacterial infections and co-infections, immunisation against bacterial diseases is emerging as a promising management strategy in areas of high endemicity. Systemic infections with typhoidal and non-typhoidal *Salmonella enterica* rank among the most common bloodborne infections in Sub-Saharan Africa. Designing affordable, safe vaccines that elicit durable responses against a range of *S.* enterica serovars is challenging and would benefit from a detailed understanding of the effects that different vaccine candidates exert on the within-host dynamics of *Salmonella*. In this study, I use mechanistic mathematical models applied to ITS-based data comparing the effects of a live- and killed- *S.* Typhimurium vaccine on the in-host bacterial dynamics following a secondary challenge. Inferences show that the superior performance of the live vaccine in controlling both bacteraemia and intra-organ bacterial proliferation over a longer time period is attributable to an early inoculum inactivation, early bacteriostatic effects in the liver and cidal in the spleen, followed by a later period of bacterial trapping in the organs and control of intra-organ growth. These effects all align with the predicted effects of a significant IgG2 response which is induced in mice immunised with the live but not the killed vaccine.

4.1. Introduction

In the 21st century, bacterial diseases persist as leading causes of death and disability worldwide, particularly in low-resource settings, disproportionately affecting the immunocompromised, elderly and under five population (Gordon *et al*., 2008). Both typhoidal and non-typhoidal *Salmonella* are major causative agents of bloodstream infections and large contributors to morbidity and mortality in Sub-Saharan Africa (Feasey *et al*., 2015; Marks *et al*., 2017), accounting for almost 40% of community-acquired bloodstream infections in the region (Uche, MacLennan and Saul, 2017).

With multidrug-resistant *Salmonella* infections on the rise worldwide (Yan *et al.,* 2016; Feasey *et al.*, 2015), the WHO has advised the mass immunisation of individuals living in areas of invasive salmonellosis endemicity (Breiman *et al.,* 2012). Although prevention strategies have been an active area of pharmaceutical research and development since the 1890's, three vaccines against *S.* Typhi have been licensed to-date (Guzman *et al.*, 2006), and no vaccines are available against other *Salmonella* serovars responsible for paratyphoidal and non-typhoidal salmonellosis (Coward *et al.,* 2014). The increasing rates paratyphoidal and non-typhoidal invasive salmonellosis prevalence in endemic countries (Ochiai *et al*., 2005), combined with the >20% fatality rate of non-typhoidal salmonellosis amongst children (Tapia *et al.*, 2015) renders the need for vaccines conferring protection against a wider range of *Salmonella* serovars pressing.

With vaccination emerging as a promising solution to the aggravating problem of antimicrobial resistance, efforts are being made to accelerate the research and development (R&D) of new vaccines. High-throughput screening of large libraries of vaccine candidate targets (Davies *et al.*, 2012; Guo *et al.*, 2015; Itoh *et al.*, 2019), and a detailed understanding of how pathogens interact with their host to colonise tissues, disperse and proliferate (Hadjichrysanthou *et al.*, 2016; Coward *et al.*, 2014; Orwa, Mbogo and Luboobi, 2018; Handel *et al.*, 2018; Davies *et al.*, 2019) are gradually being recognised as major accelerators of R&D.

While the *en masse* screening of vaccine candidates has been enabled by biomolecular technological advances and immunological research (Broach and Thorner, 1996), novel insights into how pathogens replicate, move and die in their hosts are being generated

through the increasing use of mathematical modelling as an adjunct to conventional data analysis (Restif, Thakar and Harvill, 2009). Studies using sophisticated experimental techniques are revealing finer and finer details at the host-pathogen interface (Grant *et al.*, 2008; Helaine *et al.*, 2010; Walters *et al.*, 2012; Claudi *et al.*, 2014; Korem *et al.*, 2016). However, without the application of mathematical models, the output of such studies remains conventionally limited to measurements of net changes in pathogen numbers (Walters *et al.*, 2012) or, at best, bacterial replication events (Helaine *et al.*, 2010; Claudi *et al.*, 2012).

Mechanistic insights into the within-host dynamics of pathogens benefits vaccine R&D in three ways. First, it has been historically recognised that the most effective vaccines are those that induce an immune response closely mimicking that of a natural infection. A detailed understanding of how, when and where natural immunity controls an infection can inspire the design of vaccines that either mimic or act in synergy with natural host defences (Garon and Orenstein, 2015). Second, by characterising the mode of action of candidate vaccines prior to clinical trials, it is possible to predict potential shortcomings in terms of their performance and safety profile in populations with immunodeficiencies, co-infections or profiles that correlate with an altered immunological status (e.g. elderly individuals, young children or pregnant women). Third, given that vaccines exert pleiotropic effects on the host's immune system, the quantitative characterisation of each of these effects on the demographic events that shape the size of bacterial subpopulations in the tissues can help determine the predominant effect(s) of any given vaccine candidate. For example, if a vaccine induces mainly bactericidal effects, it is possible to investigate this mechanism further and pinpoint the component(s) of the host immune response that is responsible for this effect. This host immunity response can then be combined with other favourable responses to create a potent and safe vaccine.

In this thesis, I already addressed the first point in Chapter 3, showing differences between the liver and spleen microenvironments in terms of their ability to control the proliferation of *S.* Typhimurium over time; these differences are attributable to specific host immunity factors which need to be further investigated. In Chapter 4, I aim to address the second point by quantitatively characterising the effect of two vaccine formulations, a live-attenuated (LV) and an inactivated, whole-cell vaccine (KV) on the within-host dynamics of a secondary intravenous challenge with *S.* Typhimurium.

A live-attenuated *S.* Typhimurium vaccine was first constructed by Hoiseth and Stocker (Hoiseth and Stocker, 1981) by using deletion mutations in the *aroA* locus, thus rendering *Salmonella* growth strictly conditional on the presence of aromatic compounds, which are not found in mammalian tissues. It has remained one of the most promising candidates since (Table 4.1), the only caveat being that the safety profile of live vaccines amongst immunocompromised individuals remains a long-standing challenge, as on the one hand, knowledge about their effects in these populations is lacking (Pirofski and Casadevall, 1998), and on the other, a large proportion of individuals living in *Salmonella*-endemic areas are infected with one or more immunomodulatory diseases. In a study by Izhar *et al.* (Izhar *et al.*, 1990) it was reported that moderately immunocompromised mice infected with the Δ*aroA Salmonella* vaccine strain showed no greater susceptibility to infection than immunocompetent mice; this finding validates that the virulence attenuation in *aroA* mutants is not attributable to effect(s) at the host immunity-pathogen interface, and makes further exploration of the effects of this vaccine worth pursuing.

Live-attenuated vaccines have represented an effective alternative to killed vaccines, which were among the first to be tested *in vivo* due to their perceived good safety profile (Cronly-Dillon, 1972). Evidence collected over the years (Cronly-Dillon, 1972; Angerman and Esenstein, 1980; Coward *et al*., 2014) showed that any protection conferred by killed-vaccines, both acetone- and heat-killed, is humoural and not durable.

Study	Strain of Immunising *S.* Tm	Processing of *S.* Tm strain before immunisation	Route of immunisation	Characterised effects of immunisation
Cronly-Dillon, 1972	1566	1. Heat-killed *S.*Tm	Intraperitoneal	At 0.5 months post-immunisation, no protection was offered in oral challenges, and equal protection in intraperitoneal challenges for preparations 1 and 2.
		2. Acetone-killed *S.*Tm		
Angerman & Eseinstein, 1980	W118-2	1. Low-dose live *S.*Tm	Intraperitoneal	At ~3 months post-immunisation, the percentage of mice surviving a secondary challenge are 50%, 0%, 20% and 0% for preparations 1-4 respectively.
		2. Acetone-killed *S.*Tm		
		3. Ribosomal extract		
		4. LPS extract		
Harrison *et al.*, 1997	SL3261	1. Acetone/Heat-killed *S.*Tm	Subcutaneous	At ~2 months post-immunisation, live Δ*Aro*A S.Tm induced a Th1 cellular response with higher anti-LPS IgG2a levels than both killed vaccines. The live Δ*Aro*A *S.*Tm protection was equally durable across all immunisation routes.
		2. Live Δ*Aro*A *S.*Tm	Subcutaneous	
			Intraperitoneal	
			Oral	
Coward *et al.*, 2014	SL3261	1. Acetone-killed *S.*Tm	Subcutaneous	At ~1.5 months post-immunisation live Δ*Aro*A *S.*Tm vaccine rapidly inactivated ~40% of challenge inoculum, induced bacteriostasis and a delayed bactericidal T-cell response.
		2. Live Δ*AroA S.*Tm		
Lim *et al.*, 2014	SL1344	Live Δ*AroA S.*Tm	Oral	At ~1.5 months post-immunisation, invasion of the secondary challenge to the Payer's patches was reduced, but not abolished.
Erova *et al.*, 2016	14028	1. Live *ΔlppAB S.*Tm	Oral	At ~1 month post-immunisation, both preparations provide complete protection against a secondary challenge. Live Δ*lpp*AB Δ*msb*B *S.*Tm-immunised mice produced higher levels of faecal IgA and IL-6, while live Δ*lpp*AB *S.*Tm-immunised mice produced higher levels of IL-17A and TNF-α.
		2. Live *ΔlppAB ΔmsbB S.*Tm		
Diard *et al.*, 2019	SB300	Peracitic acid killed *S.*Tm	Oral	At ~1 month post-immunisation, immunised mice were incompletely protectected due to the emergence of *S.*Tm mutants with short O-antigens, which were not bound by vaccine-induced IgA.

Table 4.1: Review of *in vivo* studies on the effects of immunisation preparations on *S.* Typhimurium dynamics.

In this chapter, I show how using data from experiments using ITS of *S.* Typhimurium in immunised mice, in conjunction with bespoke mechanistic models can suggest novel biological insights on the effect of different vaccines (Coward *et al.*, 2014). I focus on data from an ITS-based study by Coward *et al*., (2014), to study the effects of a live-attenuated and an acetone-killed vaccine on the within-host dynamics of *S.* Typhimurium post-immunisation. Based on previous work on the course of systemic *Salmonella* infections in the naïve host (Chapter 3), I develop a mechanistic mathematical model which assumes free bacterial movement between the blood and the two main reservoirs of infection: the liver and spleen (Grant *et al.*, 2008), with concomitant bacterial replication and killing occurring

in the latter. In the differentially immunised groups, I quantify the differential bacteriostatic and cidal effects of the two vaccine formulations and their ability to control bacterial spill-over from the organs to the blood. This work builds on preliminary modelling work in Coward *et al*. (Coward *et al.* 2014) who only used a simple model with no organ-to-blood migration and did not analyse data from the later phase of the infection due to computational constraints in the maximum-likelihood inference framework. My suggested model corrects the poor fit of the original model, also highlighted by Price *et al.* (2017), and extends the inferencing into the later time points to provide a complete picture of the mechanisms by which the two vaccines differentially influence the within-host dynamics of infection in a secondary *S.* Typhimurium challenge.

4.2. Methods

4.2.1. Experimental data background

Coward *et al.* (2014) collected datasets on bacterial counts of isogenic-tagged *S.* Typhimurium SL3261 strains in two groups of differentially immunised mice C57BL/6 mice intravenously challenged with *S.* Typhimurium 30 days post-immunisation. Mice used as control were not immunised (Naïve). Of the immunised groups, the first received a live-attenuated Δ*AroA* vaccine (LV), while the second received an acetone-killed vaccine (KV) subcutaneously. Mice were challenged with a ~240 CFU *S.* Typhimurium equiproportionate inoculum comprising 8 ITS at ~30 CFU per strain. At 0.5, 6, 24, 48 and 72 hours post-inoculation, 9 or 10 mice were killed, their livers and spleens harvested, blood was sampled from their tail vein, bacteria isolated and segregated by strain using qPCR. By studying the changes in ITS composition in the tissues with time and comparing strain distributions between tissues it is possible to make some preliminary inferences about the within-host dynamics of *S.* Typhimurium in the immunised murine host. Figure 4.1 illustrates the number of bacteria per ITS in each of the biological replicates (9 or 10 mice x 8 ITS/mouse = 72 or 80 total biological replicates) for each sampling time point.

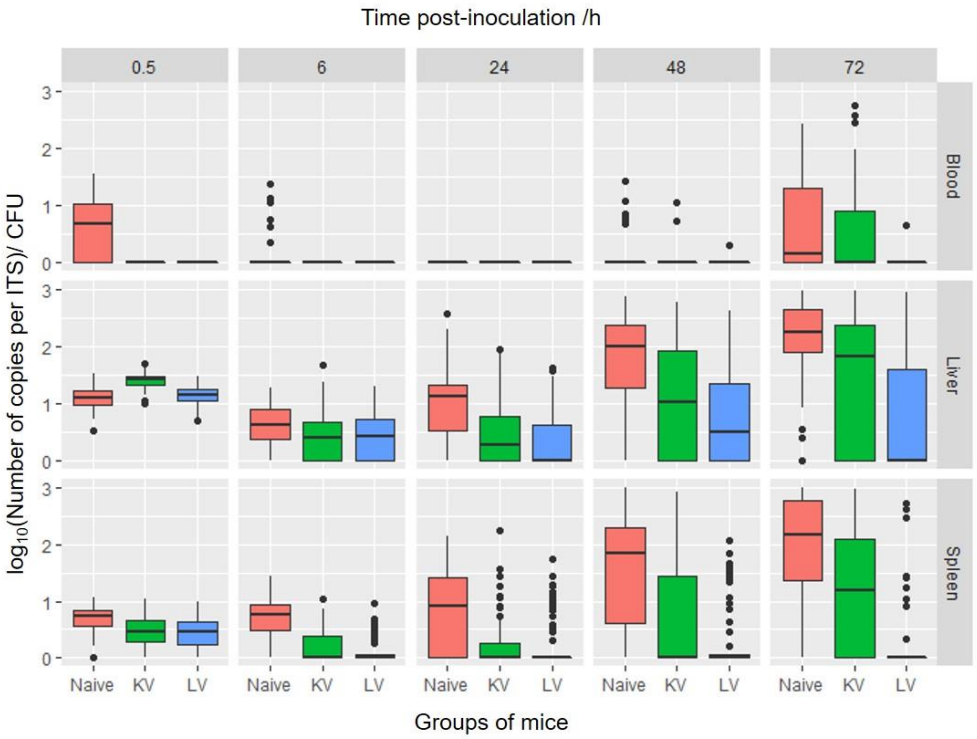


Figure 4.1: Bacterial copies per ITS in different tissues at various timepoints post-infection in the naïve, KV and LV groups of mice

Figure 4.1 shows the bacterial copies per ITS, as recovered by qPCR in the blood, liver and spleen of the naïve, KV and LV groups of mice. Across the three groups, there is negligible bacteraemia until 48 hours post-infection; in the 48-72 hour interval, the LV-group exert superior control, while the KV-group do not control bacterial migration to the blood equally well. The liver is the organ where bacteria preferentially migrate to in the first 0.5 hour post-inoculation in both the naïve and immunised groups. Both the KV and LV groups restrain bacterial numbers in the liver during the first 24 hours, with the LV group exerting a longer-lasting effect (until 72 hours). A similar trend is observed in the spleen, with the LV group controlling bacterial numbers much more effectively across the infection period, while in the KV group the control is weaning gradually after 24 hours.

4.2.2. Model structure

In this chapter, I used a general model structure, including bacterial killing and replication in the blood, liver and spleen, as well as bacterial migration between blood and the organs. The model structure can be summarised as follows (Figure 4.2), with the parameter names and respective descriptions shown in Table 4.2.

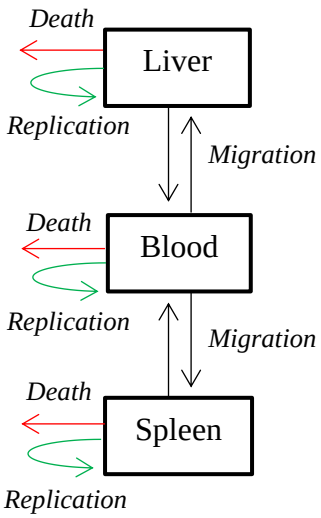


Figure 4.2: Schematic representation of the model structure applied to ITS-based data from naïve and immunised mice

	Parameter name	Description
Inter-organ migration	m1.2	Bacterial migration from the blood to the liver
	m1.3	Bacterial migration from the blood to the spleen
	m2.1	Bacterial migration from the liver to the blood
	m3.1	Bacterial migration from the spleen to the blood
Intra-organ dynamics	k1	Bacterial death in the blood
	k2	Bacterial death in the liver
	k3	Bacterial death in the spleen
	r1	Bacterial replication in the blood
	r2	Bacterial replication in the liver
	r3	Bacterial replication in the spleen

Table 4.2: Description of parameters and parameter names

4.2.3. Parameter Inference

Parameter inference for each model was performed by minimisation of the KL divergence between the set of experimentally observed moments and predicted moments given a parameter set θ, as described in section 2.5.

4.2.4. Quantification of uncertainty around parameter estimates

To quantify the uncertainty around the parameter estimates, I used a parametric bootstrap approach (outlined in 2.5 and 2.6) to generate 500 bootstrapped datasets from the best-performing model according to 3.2.3. For each bootstrapped dataset, I calculated the moments and performed inference by fitting the model and obtaining a set of best parameter estimates. For each parameter, I sorted the parameter estimates in ascending order and plotted the 95% confidence interval for that parameter.

4.3. Results

4.3.1. Early inoculum inactivation

As shown by the raw data in Figure 4.1, a large proportion of bacteria is killed or inactivated in the blood of the LV mice during the first 30 minutes post-inoculation Using the data from the early sampling time point (0.5 hours post-inoculation), it is possible to estimate the number of bacteria that actively migrate to and die/replicate in the organs, and only consider those bacteria seeding the tissues as the active inoculum dose. My inference suggests that in the LV group, over 1/3 of the inoculum is inactivated (Figure 4.3) in the first 30 minutes of infection, unlike the KV and naïve group where the inferred inoculum size post-inoculation is ~ 30 CFU per ITS, equal to the administered inoculum dose.

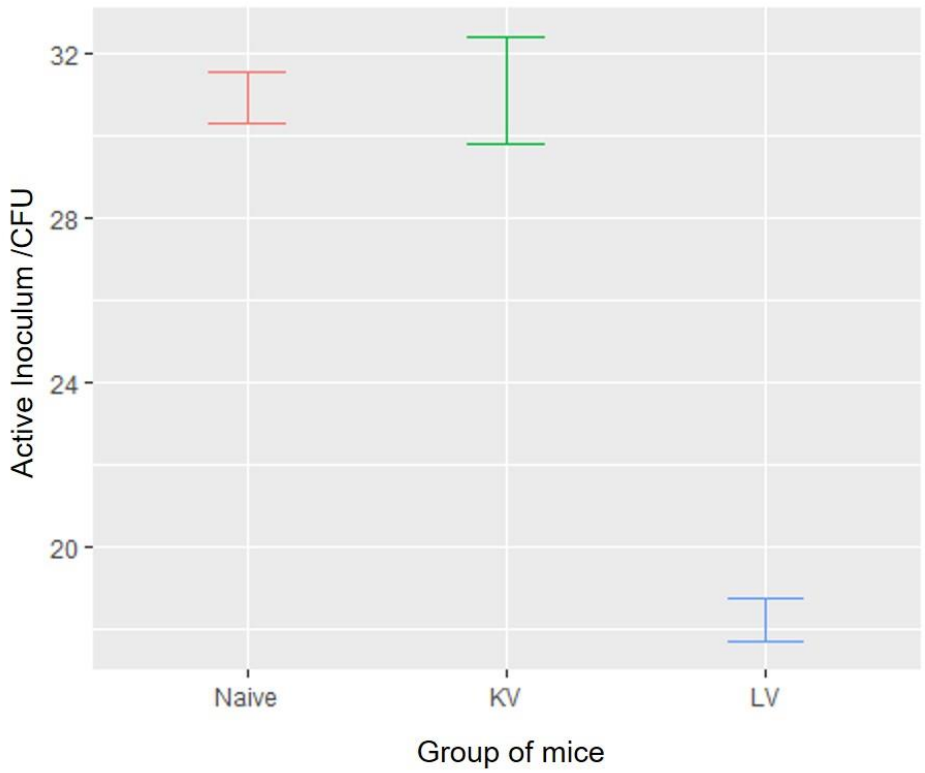


Figure 4.3: Bar plots of 95% confidence intervals for the active inoculum estimate in naïve, KV and LV mice at 0.5 hours after inoculation, estimated using parametric bootstrap (n = 500).

4.3.2. Comparison of intra-organ dynamics between KV- and LV- immunised mice

This section compares the bacterial killing and replication rates in the liver and spleen of naïve, KV- and LV- immunised mice. Inferences have been made using the model presented in 4.2.2, and results are presented following the phases of infection, in alignment with my previous analysis in Chapter 3.

- **0 to 6 hours**

Mice immunised with either the whole cell or the live-attenuated vaccines show no signs of bacteraemia within the first 6 hours of infection, unlike the naïve group as shown in Figure 4.1. In line with Coward *et al.* (Coward *et al*., 2014), my modelling work estimates that upon inoculation, a fraction of the original bacterial population is rapidly inactivated in the blood (Figure 4.3), while the remaining bacteria are asymmetrically distributed between the liver and spleen (Figure 4.4). Of the active inoculum, the majority of bacteria colonise the liver in both the KV- and LV- immunised groups in the first 30 minutes post-inoculation, similar to the naïve group (Figure 4.4).

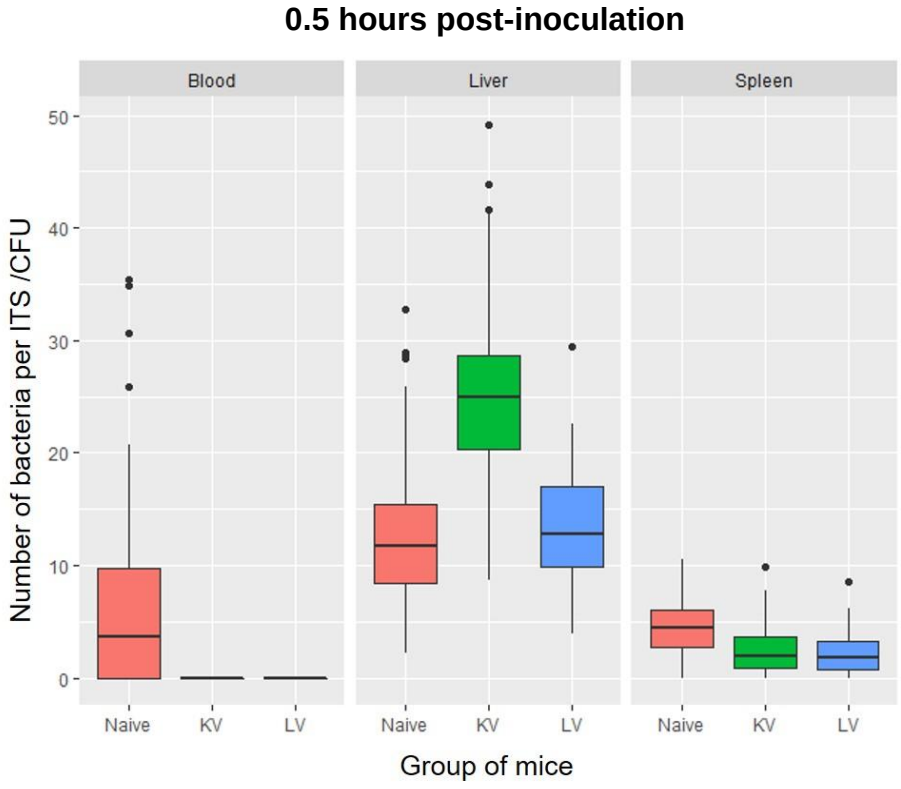


Figure 4.4: Boxplots showing the experimentally observed number of bacteria per ITS in the blood, liver and spleen for the naïve, KV- and LV- groups of mice.

Following organ colonisation, bacterial numbers decline sharply within the first few hours. In the liver, bacterial numbers decline due to high bactericidal activity in the naïve mice, while in the LV and KV groups the rates of bacterial death and replication are lower, suggesting that the vaccines elicit a bacteriostatic effect (lower r2) which, in turn, renders bacteria more recalcitrant to killing (Figure 4.5).

In the spleen, the numbers of bacteria in the naïve group increase during the first 6 hours post-inoculation (positive net growth), while in both the KV and LV mice bacterial numbers decline (ellipses in dashed lines in Figure 4.6) due to the higher bactericidal activity in both immunised groups (Figure 4.5). For the two immunised groups, bacterial killing and replication rates in the spleen do not differ substantially.

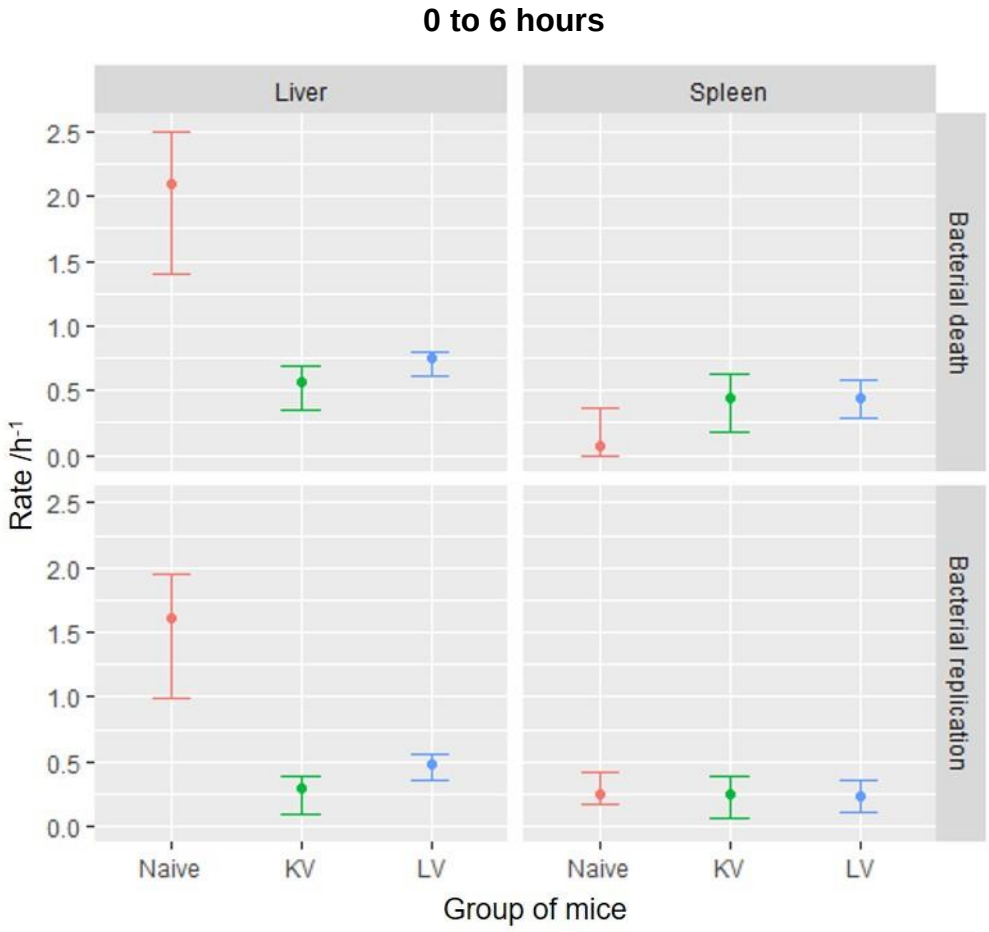


Figure 4.5: Bar plots of 95% confidence intervals for the rates of bacterial replication and death in naïve, KV and LV mice from 0 to 6 hours after inoculation, estimated using parametric bootstrap ($n = 500$). The dots represent estimated value for each parameter.

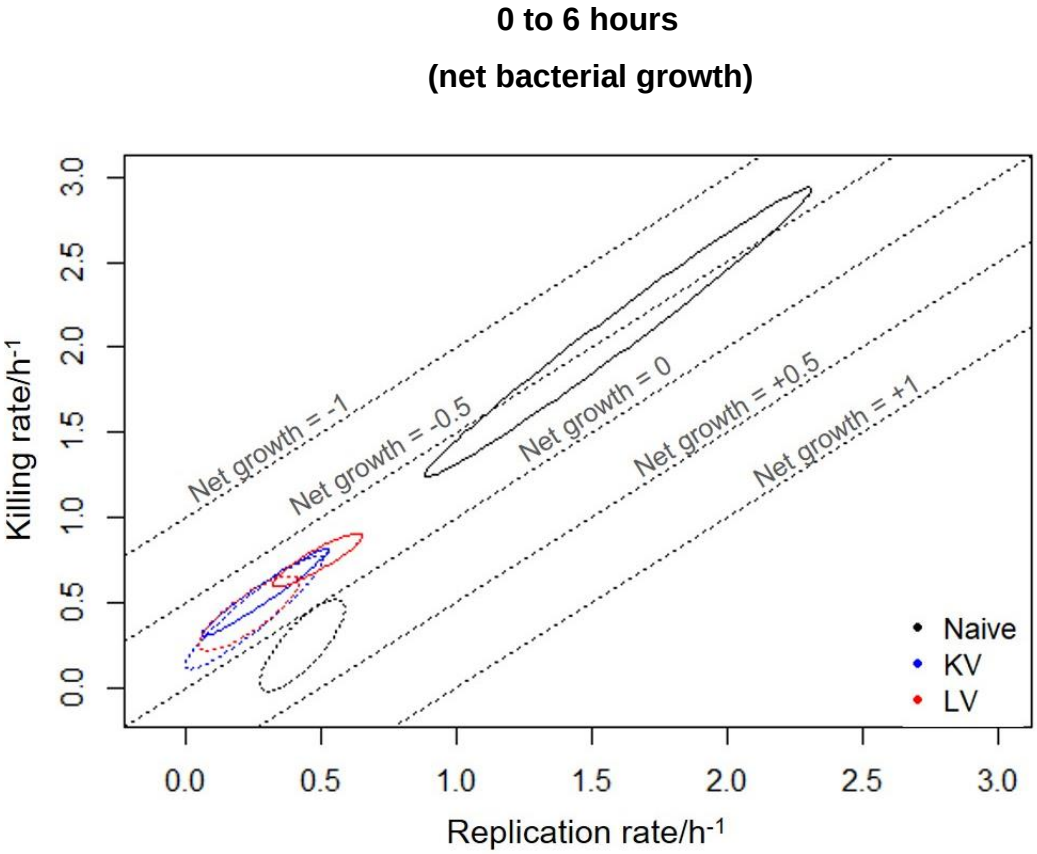


Figure 4.6: Bivariate 95% confidence regions for the rates of bacterial replication and death in naïve, KV and LV mice from 0 to 6 hours after inoculation, estimated using parametric bootstrap ($n = 500$)

The black, blue and red ellipses illustrate the bivariate 95% confidence region for replication and killing rate in naïve, KV and LV groups of mice respectively, with the solid lines representing the liver and the dashed lines the spleen. The large length of the major axis of the ellipse represents the high correlation between replication and killing parameter estimates, which accounts for the large confidence intervals seen in figure 4.5. Diagonal grid lines indicate net growth rates (replication rate-killing rate). While bacteria proliferate in the spleens of naïve mice, there is net bacterial death in the spleens of both the KV and LV mice, secondary to enhanced bactericidal activity, as shown in Figure 4.5. Net bacterial death occurs in the liver of naïve, KV and LV mice alike.

- **6 to 24 hours**

During the second phase of infection, no signs of bacteraemia are present in the immunised groups. Inference suggests that the patterns of intra-organ dynamics remain similar to the first phase, with estimates of bacterial replication and killing in the liver lower for both immunised groups; this reflects a persisting bacteriostatic effect (Figure 4.7). In the spleen, the rates of bacterial replication and death remain similar for the immunised groups, while the bactericidal action in the spleen of naïve mice increases leading to a reduction in the net growth rate (Figure 4.8).

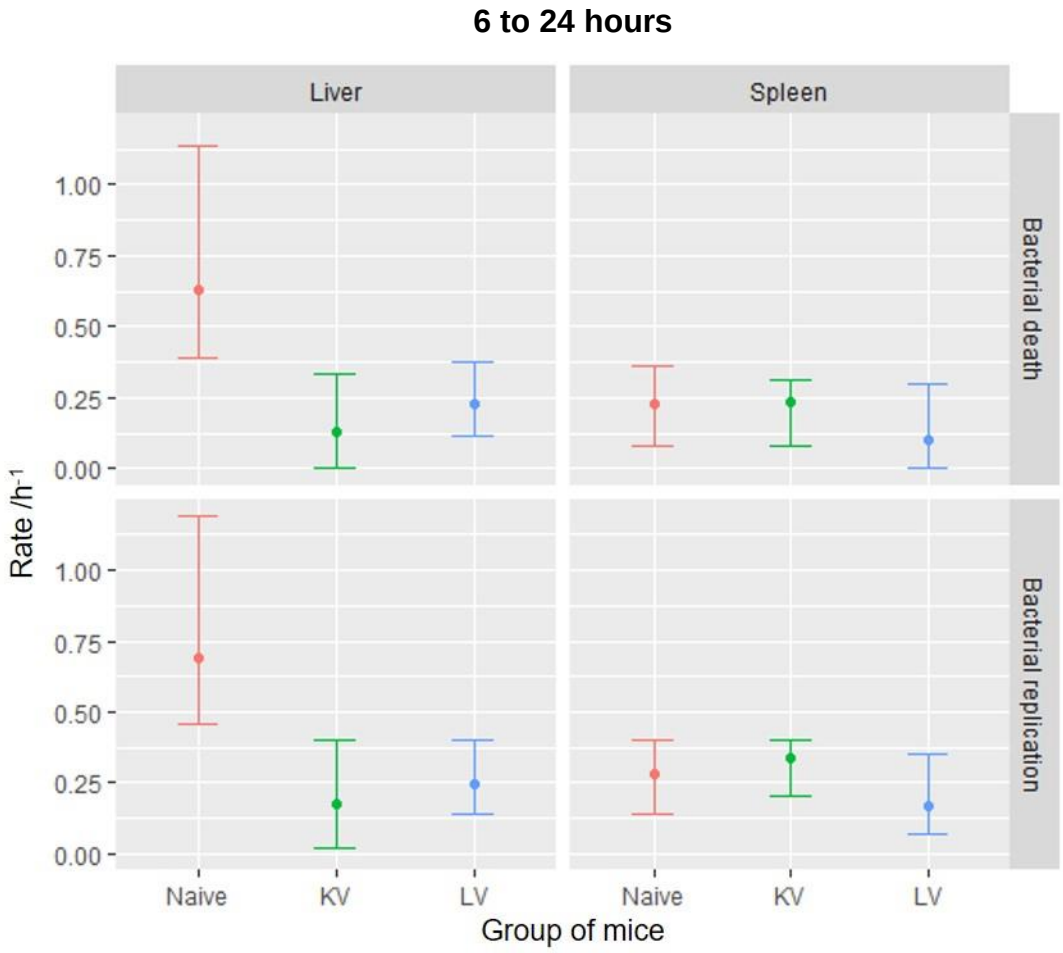


Figure 4.7: Bar plots of 95% confidence intervals for the rates of bacterial replication and death in naïve, KV and LV mice from 6 to 24 hours after inoculation, estimated using parametric bootstrap ($n = 500$). The dots represent the estimated value for each parameter.

Bacterial numbers increase in the liver and spleen of both immunised groups, albeit at a slower rate compared to the naïve group; intra-organ net growth rate in immunised mice is lower than the net growth rate in naïve mice (Figure 4.8).

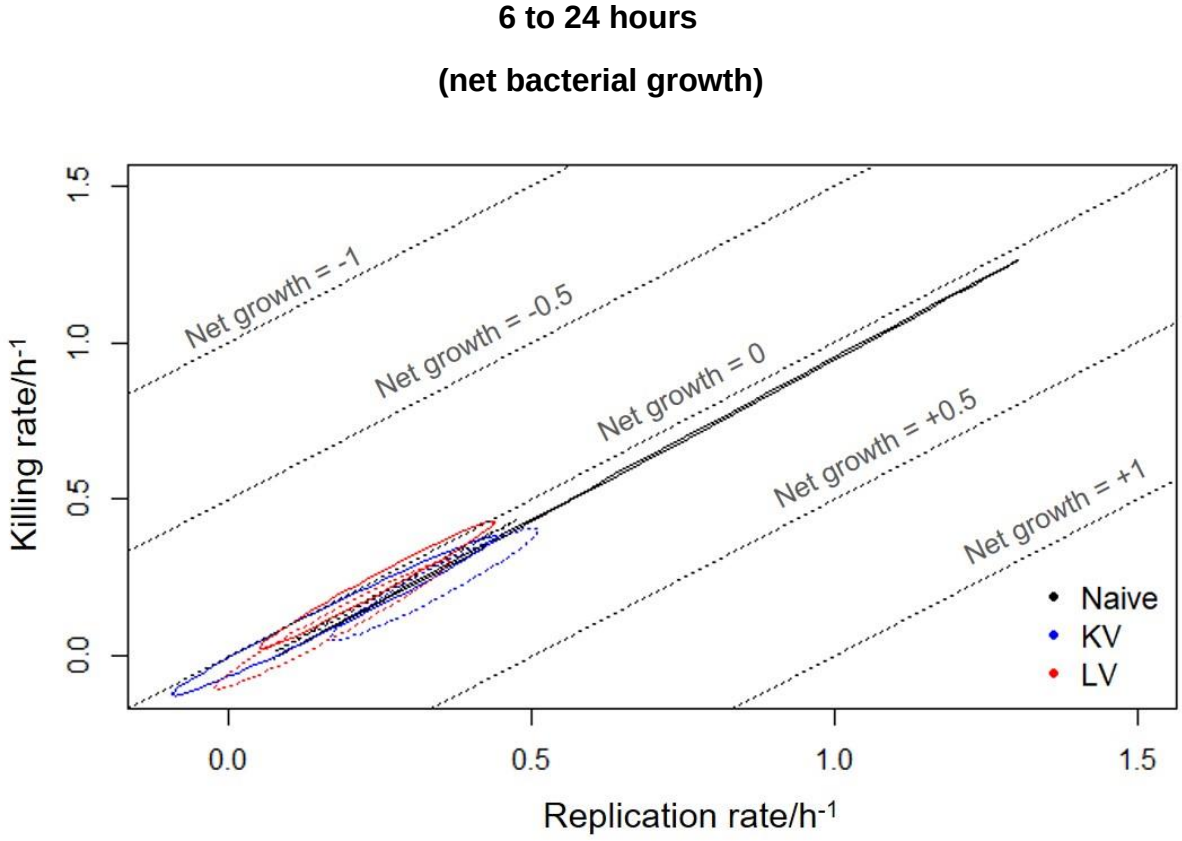


Figure 4.8: Bivariate 95% confidence regions for the rates of bacterial replication and death in naïve, KV and LV mice from 6 to 24 hours after inoculation, estimated using parametric bootstrap ($n = 500$).

The black, blue and red ellipses illustrate the bivariate 95% confidence region for replication and killing rate in naïve, KV and LV groups of mice respectively, with the solid line representing the liver and the dashed line the spleen. The large length of the major axis of the ellipse represents the high correlation between replication and killing parameter estimates, which accounts for the large confidence intervals seen in figure 4.7. Diagonal grid lines indicate net growth rates (replication rate-killing rate). The ellipsoid confidence regions for all groups are congregated at the bottom left-hand side of the graph, indicating marginal net growth for all groups, with the highest values corresponding to the KV and naïve groups.

- **24 to 72 hours**

During the third phase of infection, signs of bacteraemia first appear in the naïve and KV group (by 48 hours), while bacteraemia is delayed in the LV group with bacteria detected in the blood only after the 48-hour mark. In terms of intra-organ dynamics, it appears that in the spleen the rates of replication and death are similar across the three groups, as indicated by the overlapping confidence intervals between 24 and 48 hours. In the 48-72 hour interval, bacterial replication accelerates in the immunised groups, but not in the naïve group. In the liver, bacterial replication and death are consistently the highest in the naïve group (Figure 4.9).

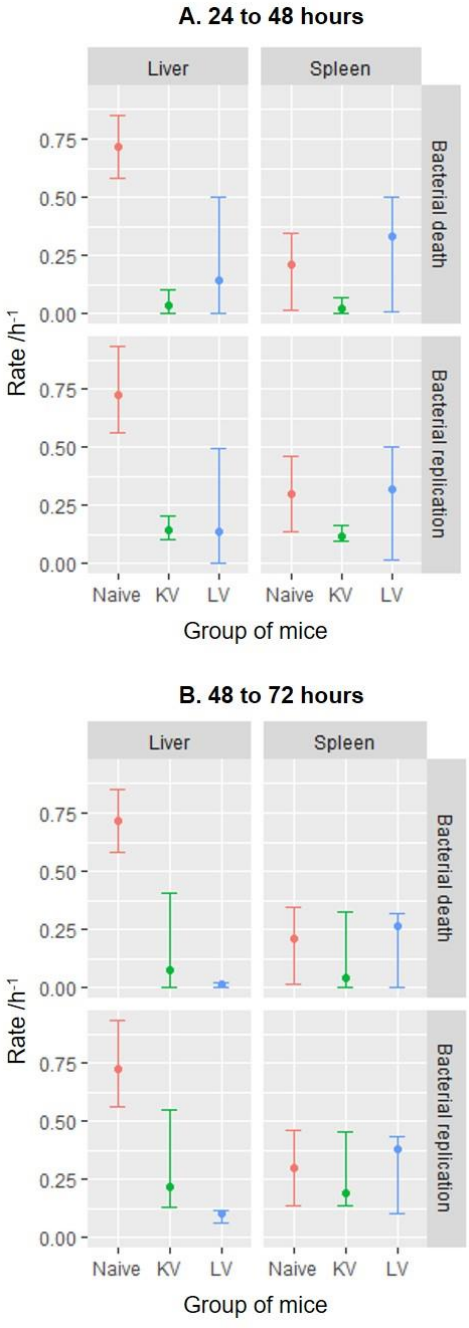


Figure 4.9: Bar plots of 95% confidence intervals for the rates of bacterial replication and death in naïve, KV and LV mice from 24 to 74 hours after inoculation, estimated using parametric bootstrap ($n = 500$). The dots represent the estimated value for each parameter.

Panels A and B offer a refined account of the intra-organ dynamics during the third phase of infection, as the dynamics in the immunised groups change more readily than in the naïve group, as evidenced by the differentially timed appearance of bacteraemia in the KV and LV groups.

In the liver, the live- and killed- vaccine both induce bacteriostatic effects, with concomitant lower bacterial death (panels A and B). In the spleen, there is evidence of bacteriostasis in the KV, but not the LV group (panels A and B). These conclusions are not definitive and only suggested by the point estimates (dots), as the 95% confidence intervals for the parameter estimates for bacterial death and replication in the spleen are overlapping across the three groups.

In the LV group net bacterial growth in both the liver and the spleen is the lowest between 24 to 48 hours, indicating a better control of the infection both in terms of bacteraemia and in terms of intra-organ proliferation (Figure 4.10 (A)). However, between 48 and 72 hours, the LV group no longer retains its ability to control intra-organ bacterial growth, as shown in Figure 4.10 (B). In contrast, in the KV group positive net bacterial growth starts earlier in both the liver and the spleen (Figure 4.10 (A)) and remains at the same level throughout the 24 hours (Figure 4.10. (B)).

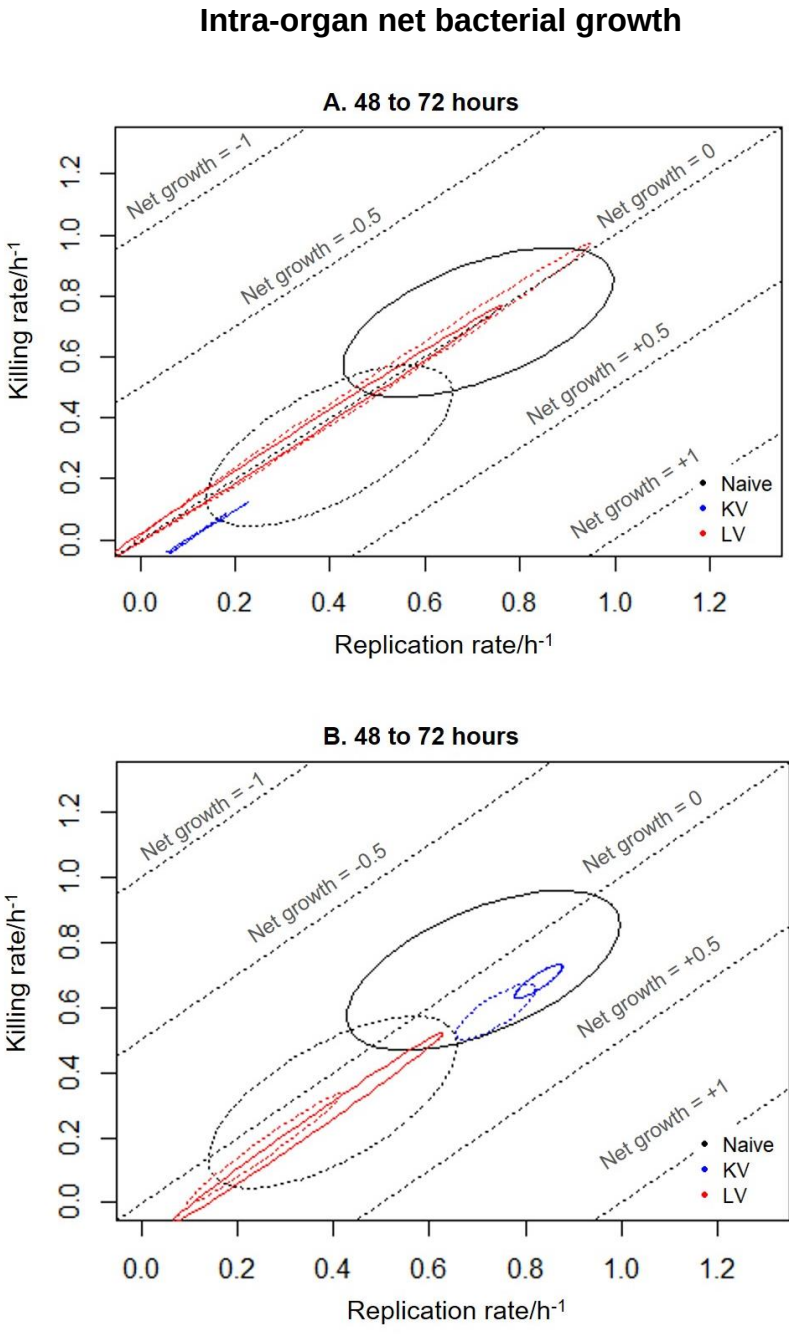


Figure 4.10: Bivariate 95% confidence regions for the rates of bacterial replication and death in naïve, KV and LV mice from 0 to 6 hours after inoculation, estimated using parametric bootstrap ($n = 500$)

The black, blue and red ellipses illustrate the bivariate 95% confidence regions for replication and killing rate in naïve, KV and LV groups of mice respectively, with the solid line representing the liver and the dashed line the spleen. The large length of the major axis of the ellipse represents the high correlation between replication and killing parameter estimates, which accounts for the large confidence intervals seen in figure 4.7. Diagonal grid lines indicate net growth rates (replication rate - killing rate).

Panels A and B illustrate the comparative net growth in the three groups between 24-48 hours, and 48-72 hours post- infection respectively. While positive net growth starts from 24 hours post-infection in both the KV and naïve group, in the LV group it only starts after 48 hours, suggesting that the live-attenuated vaccine induces longer-term control of bacterial proliferation in both the liver and spleen.

4.4. Discussion

Following my analysis in Chapter 3, in this chapter I applied a likelihood-free inference framework to published experimental datasets by Coward *et al*., 2014 to quantify and compare the within-host dynamics of intravenously-acquired *S.* Typhimurium infections in mice immunised with either an live-attenuated or an acetone-killed vaccine. Qualitative patterns in the data suggest that between the two vaccine formulations, the live-attenuated vaccine exerts superior control both in terms of lower intra-organ bacterial proliferation and delay of bacteraemia. Combining the data with mathematical inference, I showed that a combination of mechanisms underpin the performance of the live-attenuated vaccine in controlling intra-organ bacterial growth and bacteraemia; in particular: 1) a quick early inactivation of 1/3 of the administered inoculum, 2) bacterial trapping in the liver and spleen, 3) early bacteriostatic effects in the liver and bactericidal effects in the spleen, 4) longer-term control of intra-organ bacterial proliferation in the late phase (24-72 hours), 5) long-term trapping of bacteria in the organs until 72 hours post-infection. On the contrary, the whole-cell killed vaccine exerts more moderate control both in terms of intra-organ bacterial proliferation and bacteraemia. In the early phase of the infection, bacteriostatic mechanisms prevail in the liver, combined with moderate cidal activity in the spleen. Similar to the LV group, bacteraemia is initially well-controlled, but by 48 hours post-infection bacteria are detectable in the blood. Finally, in the last phase of the infection, unlike the LV group, positive net bacterial growth occurs in both the liver and spleen of the KV group.

The experimental observations by Coward *et al.* (2014) recapitulate patterns from an earlier study by Izhar *et al.* (Izhar *et al.*, 1990), who previously compared the *in vivo* effects of live-attenuated and whole-cell killed vaccines by relying on quantifying the inoculum size required to kill 50% of the animal host cohort (LD_{50}), a much less granular approach compared to the integrated data-based model approach followed here. In this early study, immunisation with an *aroA* mutant strain (live-attenuated vaccine) conferred the most robust protective response, with an LD_{50} of 3 logs higher than that for the killed vaccine. However, the immunological mechanisms underpinning this effect and the time frame of their action has so far remained unclear largely due to a lack of tools to investigate the dynamic components of the host-pathogen interactions: where are the cidal hotspots? where is bacterial replication slowed down? when do bacteria die and when does replication take off? do inter-organ bacterial populations mix, and if so, when and where?

In this modelling study, I have addressed these questions for the 72-hour interval following the secondary intravenous challenge. To contextualise these findings and evaluate their impact, it is important to consider the different pathways of immunity that underpin the effects of the two vaccine formulations, and relate them to the patterns of intra-organ control inferenced from the data. Live vaccines based on mutations in the aromatic pathway have been shown to elicit both humoural and cellular immunity, both of which are recognised as important in the control of *S.* Typhimurium infections in the murine host (Mastroeni, Villarreal-Ramos and Hormaeche; 1993). Despite the ability of live-vaccines to induce a cellular response, a comparison of bacterial loads between T-cell positive and T-cell negative mice showed that the T-cell response elicited by the live-vaccine only became important after the 72-hour mark; as a result, the differential effects elicited by the two vaccine preparations in the 72-hour interval examined here are attributable to differences in humoural responses.

Intravenous vaccination with a live *S.* Typhimurium vaccine has been found to elicit a a greater IgG2a and IgG2b antibody responses against a secondary *S.* Typhimurium challenge, compared to the killed vaccine (Harrison *et al*., 1997). Later studies in the murine host showed that IgG2a antibodies bind FcγRI receptors and play a key role in enhancing the antimicrobial action of reactive oxygen intermediates (Goh *et al*., 2011; Uppington *et al*., 2006). Furthermore, previous mechanistic modelling studies based on *in vitro* data on antibody treatment and infection of murine phagocytic cells showed that bacterial opsonisation with IgG2 may elicit a non-replicating bacterial phenotype, which is observed as initially low, but subsequently rising intracellular bacterial counts, given that bacteria are predicted to regain their proliferative capacities in these permissive host cells (Restif *et al*., 2013). IgG2a has been pinpointed as the main antibody subclass whose levels appear to be significantly higher in the LV, as opposed to the KV and naïve groups. Indeed, the inferred early inoculum inactivation in the LV group agrees with the findings by Restif *et al.* (Restif *et al*., 2013).

Between 0 and 6 hours, my estimates for the intra-organ dynamics differ from those reported by Coward *et al.* (Coward *et al.*, 2014), who used a simplified model without organ-to-blood bacterial migration for inference in the naïve group. As shown in Chapter 3, the fit of this simplified model to the data is poor, and it is greatly improved by adding non-zero parameters for organ-to-blood migration (m21, m31) in the early stages of infection; the new

extended model leads to higher estimates for bactericidal activity in the liver of naïve mice (Chapter 3). By comparison to the naïve group, both the KV- and LV- groups exert bacteriostatic effects in the liver during the first 6 hours of infection, and bactericidal effects in the spleen, which can be explained by the fact that LV-enhance IgG-opsonization of *Salmonella* enhances the capacity of dendritic cells to deploy lysosomal-dependent degradation of the internalised bacteria, and to present antigenic peptide–MHC complexes in the spleen (Riquelme, Bueno and Kalergis, 2012). The relative abundance of dendritic cell populations in the spleen could explain the bactericidal activity inferred in the splenic compartment during the early phase of the infection.

In the late phase of infection (24-48 hours) the dynamics in the vaccinated groups differ, with the LV-group exerting superior control on bacterial proliferation than the KV-group during the 24-48 hour interval in both the liver and the spleen (Figure 4.10(A)). As bacteria are mostly intracellular at this stage, it is possible that the differences in bacterial proliferation permissiveness between the two groups of immunised mice are attributable to differences in the interactions between the host cells and differentially-coated bacteria. Indeed, the main difference in antibody expression between the two vaccine preparations is the significantly higher levels of IgG2 in the LV group (Harrison *et al*., 1997). A later *in vitro* study used infections with *S.* Typhimurium coated with humanised antibody on human phagocytes and reported that IgG2-mediated phagocytosis was preferentially dependent on FcγRIIA than FcγRI receptors, while for all IgG subclasses (IgG1, IgG3, IgG4) bacterial internalisation depended more on FcγRI (Goh *et al*., 2011). Given the well-established heterogeneity in expression of FcγR receptor subtypes on the surface of different leukocyte types (Rosales, 2017; Patel, Roberts and Barb, 2019), it is possible that *Salmonellae* in the two immunised groups are preferentially phagocytosed by host cells with different permissiveness to intracellular proliferation.

Overall, in this chapter I comparatively characterised the effects of a live-attenuated and an acetone-killed *S.* Typhimurium vaccine on the demographic events that underlie bacterial proliferation in the tissues during a 72-hour time interval following a secondary intravenous challenge. The early inoculum inactivation and superior control of intra-organ bacterial proliferation elicited by the LV-vaccine are both compatible with previously studied roles of IgG2, the IgG isotype whose expression is significantly higher in mice vaccinated with a live vaccine. Further *in vivo* research at the level of the single bacterium is needed to study

the effects of different vaccines on the association between different host cell types and bacteria. A blueprint for this kind of studies was recently provided by Kanvatirth *et al.* (Kanvatirth *et al.*, 2020) who studied the host cell types that *S.* Typhimurium preferentially associates with before and after antibiotic treatment. Combining the insight from such studies can help determine whether basis of the differential vaccine success in infection control is indeed due to an effect on bacterial-host cell associations in the immunised hosts.

5.

S. Typhimurium dynamics in the antibiotic-treated host

Summary Points:

- The well-documented phenotypic heterogeneity of *in vivo* bacterial populations influences the impact different antibiotic therapies have on the within-host dynamics of bacteria.
- Comparisons between different antibiotic classes can reveal whether bacterial phenotypic switch in response to antibiotics is universal or antibiotic specific, and whether these effects persist after antibiotic withdrawal.
- I address this question by combining mathematical models with a single or a dual phenotype with ITS-based data using ciprofloxacin and ampicillin.
- Inferences made regarding the rates at which the dynamics of S. Typhimurium change over time indicate that while ciprofloxacin acts via cidal mechanisms, ampicillin has mixed static and cidal effects.
- Bacterial subpopulations surviving ampicillin treatment in the mesenteric lymph node (MLN) and spleen have adopted a non-replicating, antibiotic-recalcitrant phenotype.
- Surviving bacteria in the ciprofloxacin-treated group are of the growing, antibiotic-sensitive phenotype.
- Following antibiotic withdrawal in ciprofloxacin-treated mice, growth rate in the blood, liver and spleen is higher on day 1 post-antibiotic withdrawal and, subsequently decreases. In ampicillin-treated mice, the liver, spleen and MLN are repopulated by progenitors of the non-growing antibiotic-recalcitrant bacteria.

- **Declaration of Originality**

This chapter has been reproduced verbatim from a published research paper (Vlazaki et al., 2020), for which I certify that I am the sole contributor in terms of study design, execution, and manuscript writing. I also certify that I have included the relevant citations, where appropriate.

5.0. Abstract

Antibiotic therapy has drastically reduced the mortality and sequelae of bacterial infections. From naturally-occurring to chemically-synthesised, different classes of antibiotics have been successfully used without detailed knowledge of how they affect bacterial dynamics in vivo. However, a proportion of patients receiving antimicrobial therapy develop recrudescent infections post-treatment. Relapsing infections are attributable to incomplete clearance of bacterial populations following antibiotic administration; the metabolic profile of this antibiotic-recalcitrant bacterial subpopulation, the spatiotemporal context of its emergence and the variance of antibiotic-bacterial interactions *in vivo* remain unclear. Here, we develop and apply a mechanistic mathematical model to data from a study comparing the effects of ciprofloxacin and ampicillin on the within-host dynamics of *Salmonella enterica* serovar Typhimurium in murine infections. Using the inferential capacity of our model, we show that the antibiotic-recalcitrant bacteria following ampicillin, but not ciprofloxacin, treatment belong to a non-replicating phenotype. Aligning with previous studies, we independently estimate that the lymphoid tissues and spleen are important reservoirs of non-replicating bacteria. Finally, we predict that post-treatment the progenitors of the non-growing and growing bacterial populations replicate and die at different rates. Ultimately, the liver, spleen and mesenteric lymph nodes are all repopulated by progenitors of the previously non-growing phenotype in ampicillin-treated mice.

Relevant published output:

1. Vlazaki M, Rossi O, Price DJ, McLean C, Grant AJ, Mastroeni P, Restif O. 2020 A data-based mathematical modelling study to quantify the effects of ciprofloxacin and ampicillin on the within-host dynamics of Salmonella enterica during treatment and relapse. J. R. Soc. Interface 20200299.

5.1. Introduction

In the modern era, antibiotics are the cornerstone of therapeutics in bacterial disease treatment (Center for Disease Control, 2019). The term "antibiotics" refers to a heterogeneous collection of naturally-derived or chemically synthesised compounds which clear bacterial infections through various biochemical mechanisms, including inhibition of cell wall synthesis, bacterial lysis, nucleic acid and protein synthesis inhibition (reviewed in Kohanski, Dwyer and Collins, 2010). Antibiotics are traditionally classified as either bactericidal reducing bacterial load by killing bacteria, bacteriostatic acting to limit bacterial loads by keeping bacteria in a stationary growth phase, or both (Wilson and Miles, 1964; Bernatová *et al*., 2013).

The effect of different antibiotics on bacterial dynamics has been studied extensively *in vitro*, in optimally and homogeneously growing bacterial populations harvested from exponentially growing cultures (McCall *et al*., 2019). Conclusions drawn from these studies provide insights into the mode of action of different antibiotic agents on their bacterial targets; however, their clinical relevance remains unknown, as *in vivo* bacterial populations are characterised by heterogeneity in their growth and dissemination dynamics (Claudi *et al*., 2014; Bumann and Cunrath, 2017; Dewachter, Fauvart and Michiels, 2019; Haugan *et al.*, 2018). Upon successful colonisation, *in vivo* bacterial behaviour is multifactorially shaped by nutrient availability (Claudi *et al*., 2014; Shehata and Marr, 1971), host immunity (Grant *et al*., 2008; Gjini and Brito, 2016, Duneau *et al*., 2017) spatial arrangement of the infectious foci (Bottery *et al*., 2019), and host cell composition in the immediate bacterial vicinity (Kolter, Siegele and Tormo, A. 1993; Kanvatirth *et al*., 2020). The landscape of within-host bacterial dynamics is further complicated by the induction of stress responses in bacterial populations exposed to antibiotics, which alter the bacterial metabolic and transcriptional profile, slow down or stop replication, and ultimately affect the pathogen's susceptibility to the antibiotic (Kaldalu, Hauryliuk and Tenson, 2016). These adaptations may lead to incomplete clearance of the bacterial load as a result of the emergence of an antibiotic-recalcitrant, yet genetically non-resistant, bacterial subpopulation, termed "persister" (Kim and Wood, 2016; Bigger, 1944).

An important caveat of *in vitro* studies using optimally growing bacteria from exponential phase cultures is disregarding the influence of the bacterial growth rate on the therapeutic

effect of some classes of antibiotics. All clinically-used antibiotic classes exert their maximum therapeutic effect when applied to optimally growing bacteria, where cidal agents clear 99.9% of the bacterial load under clinically relevant regimens (Ferro *et al*., 2015; Nielsen, Cars and Friberg, 2011; Regoes *et al*., 2004). However, *in vitro* studies with *Escherichia coli* and *Staphylococcus aureus* have revealed that even though different antibiotic classes are homogeneously effective on exponential phase bacteria, they exert heterogeneous effects on slowly-replicating and non-replicating bacterial populations. Depending on which bacterial functions they target, different antibiotic classes are more or less active against slow- or non-replicating bacteria (Zeiler and Voigt, 1987; Fantin *et al*., 1991; McCall *et al*., 2019; Haugan, Løbner-Olesen, and Frimodt-Møller, 2018).

For example, beta-lactam antibiotics (natural and synthetic penicillin-like compounds), like ampicillin, have a range of pharmacological targets including bacterial cell wall synthesis and bacterial autolysis pathways amongst others (Ghooi and Thatte, 1995), resulting in both bacteriostatic and bactericidal properties respectively. However, bacteriostatic effects attributable to bacterial cell wall disruption may act antagonistically to the bactericidal effects due to bacterial lysis (Ghooi and Thatte, 1995). This antagonistic relationship has been experimentally observed *in vitro* as a positive correlation between the cidal effect of beta-lactams and the growth status of the targeted bacteria (Lee *et al.*, 2018; Tuomanen *et al.*, 1986) This class of antibiotics induces structural changes in the bacterial peptidoglycan cell wall leading to cell lysis via a bulge-mediated process (Yao, Kahne and Kishony, 2012) whether these morphological changes modulate bacterial replication rates and to what extent this phenotypically static effect antagonises the cidal action of the drug requires further investigation (Ocampo *et al*., 2014).

In contrast, quinolones, like ciprofloxacin, predominantly target the transcriptional machinery of the bacteria, and thus do not slow down bacterial growth at clinically relevant concentrations (Cui *et al.,* 2018). They act as DNA gyrase inhibitors, stalling bacterial chromosome replication through the formation of double-strand DNA breaks, thus preventing the replication forks from accessing the terminus (Drlica *et al*., 2007). Unlike beta-lactams whose efficacy has been linearly correlated with bacterial growth rate (Lee *et al.*, 2018; Tuomanen *et al.*, 1986), quinolones act on heterogeneously growing bacteria so long as they remain transcriptionally active (Sanders, 1988). Unlike other antibiotics the bactericidal activity of ciprofloxacin is not affected by the bacterial growth rate, both in *in*

vitro optimally-growing bacterial populations and in *in vivo* bacterial populations displaying heterogeneity in replication rates (Haugan, Løbner-Olesen and Frimodt-Møller, 2018).

Unlike bacteria in cultures, the heterogeneity of bacterial growth dynamics *in vivo* is far greater. A substantial body of literature has been dedicated to studying the *in vivo* effects of antimicrobial therapy on murine infections with *Salmonella enterica* serovar Typhimurium, especially focusing on slow- or non-replicating bacterial cells during antibiotic exposure. Despite the fundamental *in vitro* differences in the modes of action of different antibiotic classes, a large proportion of *in vivo* experimental studies involving naturally-occurring heterogeneously-replicating bacterial populations have been conducted using second-generation fluoroquinolone agents (either enrofloxacin or ciprofloxacin) in murine infections with *S.* Typhimurium (Claudi *et al*., 2014; Helaine *et al*., 2014; Kaiser *et al*., 2014). Using fluorescent-based techniques for observation of bacterial dynamics at the single-cell level, these studies have identified the lymphoid tissues and the spleen as sites of highest persister concentration (Helaine *et al*., 2014) have tried to address the replication dynamics of persisters' progenitors (Claudi *et al*., 2014; Helaine *et al*., 2014) have provided insights in the inter-organ bacterial exchange during the gastrointestinal phase of infection (Kaiser *et al*., 2014) and have studied antibiotic-driven persistence in the context of cooperative virulence (Diard *et al*., 2014) and promotion of antibiotic resistance (Bakkeren *et al*., 2019). A later study by Rossi *et al.* (Rossi *et al.*, 2017) used clinically-relevant concentrations of both a beta-lactam and a second-generation fluoroquinolone antibiotic later in the infection process (72 hours post-inoculation) to qualitatively characterise and compare the effect of the antibiotics on the within-host *Salmonella* dynamics, thus providing a blueprint for using murine infections with Isogenic Tagged Strains (ITS) to address questions on the effects of antibiotic therapy on bacterial dynamics *in vivo*.

Differences in the experimental design of these studies, including timing, dose, route of antibiotic administration, tissues examined and experimental techniques have unravelled different facets of the effects of antibiotic therapy on different tissues, as summarised in table 5.1.

Study	**Helaine *et al.*, 2014**	**Kaiser *et al.*, 2014**	**Claudi *et al.*, 2014**	**Diard *et al.*, 2014**	**Rossi *et al.*, 2017**	**Bakkeren *et al.*, 2019**
Strain of mice	C57 Bl6 and BALB/c	C57BL/6	BALB/c and 129/Sv	129 Sv/Ev	C57BL/6	129 Sv/Ev
Strain of *S.* Typhimurium	12023s	SB300, derivative of SL1344	SL1344	SB300, derivative of SL1344	SL1344	SL1344
Pre-treatment of mice with streptomycin	no	yes	yes	yes	no	yes
Route of bacterial inoculation	oral	oral	oral/ intravenous	intraperitoneal	intravenous	oral/ intravenous
Inoculum Dose (CFU)	2×10^{10}	5×10^{7}	$4 \times 10^{2} - 2 \times 10^{3}$ / $10^{7} - 2 \times 10^{10}$	5×10^{7}	10^{3}	5×10^{7}/ 10^{3}
Route of antibiotic administration	oral	intraperitoneal	intraperitoneal	oral	intraperitoneal	intraperitoneal
Antibiotic used	enrofloxacin	ciprofloxacin	enrofloxacin	ciprofloxacin	ciprofloxacin, ampicillin	ciprofloxacin/ ceftriaxone
Onset of administration of first antibiotic dose (days post-inoculation)	1	1	3	2	3	2
Cumulative daily antibiotic dose (mg/kg)	10	124	5	240	ciprofloxacin: 40 ampicillin: 300	ciprofloxacin: 120 ceftriaxone: 60
Duration of antibiotic treatment (days)	5	10	4	1	4	3
Detection of persister bacteria	yes	yes	yes	yes	*not in the scope of the study*	yes
Sites of persistence	spleen, mesenteric lymph nodes (MLN)	caecal lymph node	spleen	caecal tissues	*not in the scope of the study*	liver, spleen
Replication status of persister bacteria	Non-replicating	Slow-replicating	Replicating, Moderately-replicating, Slow-replicating	*not in the scope of the study*	*not in the scope of the study*	*not in the scope of the study*
Study of relapse phase	no	no	no	yes	yes	yes
Experimental techniques	Fluorescent-based markers coupled with live imaging microscopy	1. Fluorescent-based markers coupled with live imaging microscopy 2. ITS coupled with real time quantitative PCR	Fluorescent-based markers coupled with live imaging microscopy	1. Plating and enumeration of bacteria from tissue samples 2. Fluorescence microscopy of stained tissue sections *ex vivo*	Isogenic tagged strains coupled with Illumina sequencing	Confocal microscopy of stained tissue sections ex vivo

Table 5.1: Review of *in vivo* studies on the effects of antibiotic treatment on *S.* Typhimurium dynamics.

Despite the growing interest in characterising the effect of antibiotic treatment on different tissues, quantitative approaches to infer and separately quantify unobserved processes such as bacterial replication, death, migration and phenotypic switch remain underrepresented. Additionally, it is now commonly accepted that non-replicating bacterial subpopulations emerge as a result of antibiotic-exposure. However, the effect of different antibiotic classes, the tissues of origin of phenotypic variants, and the time point of emergence of these non-growing subpopulations still remain ambiguous. Finally, how bacterial populations recover in different tissues following antimicrobial withdrawal and how non-growing antibiotic-recalcitrant bacteria regain their replicative capacity during the relapse phase of an infection are both questions directly relevant to our understanding of bacterial disease recrudescence underpinning chronic infections.

An important first step to address these knowledge gaps was taken by Rossi *et al.* in their ITS-based study of the comparative effects of ampicillin and ciprofloxacin on *S.* Typhimurium *in vivo* (Rossi *et al.*, 2017). The novel contributions of the study included the use of both a beta-lactam and a fluoroquinolone antibiotic to directly compare their effects in the same experimental setup, the study of both the treatment and the relapse phase, the use of clinically relevant doses of antibiotics after the appearance of symptoms in line with the anticipated clinical response, instead of supramaximal early therapy, and data collection on 4 different organs involved in the systemic phase of infection, addressing not only intra- but also inter-organ dynamics. Finally, unlike previous studies carried at the single bacterial level, Rossi *et al.* used a bacterial population-level experimental technique (ITS) in a biological system with well-established heterogeneity in its dynamics (Claudi *et al*., 2014; Helaine *et al*., 2014; Kaiser *et al.*, 2014).

In this follow-up study, we complement the bacterial population-level outputs of this study by developing a family of mechanistic mathematical models to identify and quantify the experimentally unobserved processes of bacterial replication, death, inter-organ migration and phenotypic switch that collectively underpin the qualitative changes in the net bacterial number in different organs documented in Rossi *et al.* (Rossi *et al.*, 2017). We perform a simulation-based study to show that heterogeneities in bacterial phenotypes can be detectable from data collected at the level of the whole bacterial population. This methodological finding expands the scope of questions that can be addressed using well-established techniques such as ITS. Motivated by the well-documented emergence of a

persister, antibiotic-recalcitrant bacterial subpopulation upon antibiotic administration, we consider two models: (1) the single-phenotype model (SP) assumes a bacterial population with uniform dynamics, with each biological process described by a single parameter representing the average rate at which the process occurs, and (2) the dual-phenotype model (DP) assumes that the observed output of changing bacterial load in tissues is shaped by two bacterial subpopulations with different dynamics *in vivo*: one is a replicating, antibiotic-sensitive subpopulation, and the other is a non-replicating, antibiotic-recalcitrant subpopulation (Balaban, 2004). By comparing how well the best model within each of these two classes of models can recapitulate the experimentally observed dynamics, we infer that at clinically-relevant doses ampicillin, but not ciprofloxacin, induces a non-growing antibiotic-recalcitrant bacterial subpopulation.

5.2. Methods

5.2.1. Experimental Dataset Structure

This is a mathematical modelling study using experimental data by Rossi *et al.* (Rossi *et al.*, 2017) on the effect of ampicillin and ciprofloxacin on the within-host dynamics of wild-type *S.* Typhimurium in the murine host. Two groups of C57BL/6 mice were inoculated intravenously with an equiproportionate mix of 8 Isogenic Tagged Strains (ITS) of S. *Typhimurium* amounting a total of 1000 colony-forming units (CFU). The infections were allowed to develop for 72 hours, at which time a proportion of the mice from each group was culled, their organs harvested (either partially or in totality), homogenised and plated to determine the number of CFUs per organ/sample. Furthermore, genomic DNA was prepared from the recovered bacteria and these DNA samples were subjected to next generation sequencing (NGS) in order to determine the proportion at which each of the 8 ITS was present.

Of the remaining mice, group A was started with 300 mg/kg ampicillin and group C was put on a 40 mg/kg ciprofloxacin regimen intraperitoneally (i.p.) per day (maximum recommended dosage for veterinary treatment of small rodents) (Carpenter, 2012; Hawk, Leary and Morris, 2005). Antibiotic treatment was maintained for 96 hours; a proportion of mice was killed at 24-hour intervals to determine the per organ total CFU count and ITS composition. In total, data were collected at 4 equally spaced time points during the antibiotic treatment.

For the last phase of the study, the antibiotics were stopped, and the infection was allowed to relapse for 54 hours in both groups. Data were collected at 30 and 54 hours post antibiotic withdrawal.

5.2.2. Conceptual Model Structure

The liver, spleen and mesenteric lymph nodes (MLN) are well-established reservoirs for *Salmonella*e (Grant *et al*., 2008; Kaiser *et al.*, 2015) while bacteraemia is also a hallmark of systemic salmonellosis. On the basis of previous biological insight and prior modelling attempts, and acknowledging that the enteric-systemic interface has not been extensively studied in tandem, we first considered a general compartmental model allowing all possible connections between the key organs for which data were available, and with a single-phenotype bacterial population characterised by homogeneous dynamics (Figure 5.1). The parameters governing the transitions in the model, and mapping to the biological processes *in vivo* are described in table 5.2.

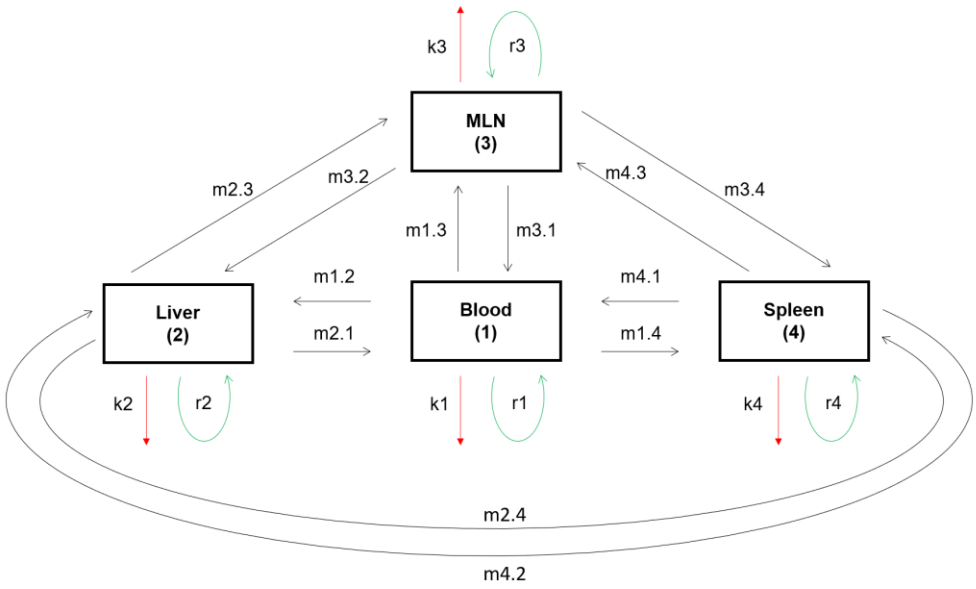


Figure 5.1: Generalised single-phenotype model with free inter-organ bacterial migration between the blood, liver, MLN and spleen, and with intra-organ bacterial death and replication in the all organs.

As an extension to the single-phenotype model and given the prolonged exposure to antibiotic-induced stressed, we proceeded to formulate a dual-phenotype model (Figure 5.2). This model assumes the emergence of a bacterial subpopulation which is non-replicating and survives antibiotic treatment (Figure 5.2).

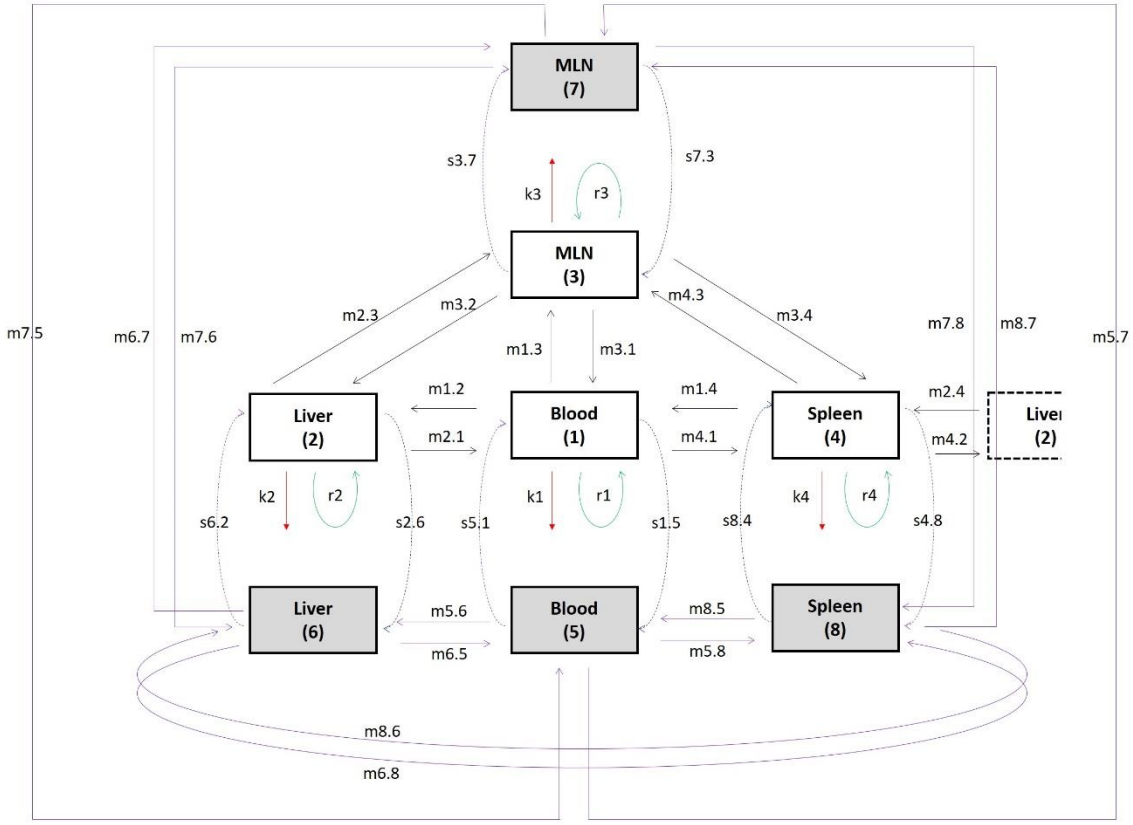


Figure 5.2: The dual-phenotype (DP) model encompasses a phenotypic switch from replicating bacteria susceptible to killing to non-growing, antibiotic-recalcitrant subpopulations in the liver, spleen and MLN, highlighted in grey.

	Parameter name	Description
Inter-organ migration (P1)	m1.2	Bacterial migration from the blood to the liver
	m1.3	Bacterial migration from the blood to the MLN
	m1.4	Bacterial migration from the blood to the spleen
	m2.1	Bacterial migration from the liver to the blood
	m3.1	Bacterial migration from the MLN to the blood
	m4.1	Bacterial migration from the spleen to the blood
	m2.3	Bacterial migration from the liver to the MLN
	m2.4	Bacterial migration from the liver to the spleen
	m3.2	Bacterial migration from the MLN to the liver
	m4.2	Bacterial migration from the spleen to the liver
	m3.4	Bacterial migration from the MLN to the spleen
	m4.3	Bacterial migration from the spleen to the MLN
Inter-organ migration (P2)	m5.6	Bacterial migration from the blood to the liver
	m5.7	Bacterial migration from the blood to the MLN
	m5.8	Bacterial migration from the blood to the spleen
	m6.5	Bacterial migration from the liver to the blood
	m7.5	Bacterial migration from the MLN to the blood
	m8.5	Bacterial migration from the spleen to the blood
	m6.7	Bacterial migration from the liver to the MLN
	m6.8	Bacterial migration from the liver to the spleen
	m7.6	Bacterial migration from the MLN to the liver
	m8.6	Bacterial migration from the spleen to the liver
	m7.8	Bacterial migration from the MLN to the spleen
	m8.7	Bacterial migration from the spleen to the MLN
Intra-organ dynamics (P1)	k1	Bacterial death in the blood
	k2	Bacterial death in the liver
	k3	Bacterial death in the MLN
	k4	Bacterial death in the spleen
	r1	Bacterial replication in the blood
	r2	Bacterial replication in the liver
	r3	Bacterial replication in the MLN
	r4	Bacterial replication in the spleen
Phenotypic switch (between P1 and P2)	s2.6	Switch from P1 to P2 in the liver
	s6.2	Switch from P2 to P1 in the liver
	s3.7	Switch from P1 to P2 in the MLN
	s7.3	Switch from P2 to P1 in the MLN
	s4.8	Switch from P1 to P2 in the spleen
	s8.4	Switch from P2 to P1 in the spleen
	s1.5	Switch from P1 to P2 in the blood
	s5.1	Switch from P2 to P1 in the blood

Table 5.2: Description of parameters used in the mechanistic models
P1 refers to the growing, antibiotic-sensitive bacteria and P2 refers to the non-growing, antibiotic-recalcitrant bacteria.

5.2.3. The modelling framework

Mathematically, these models (Figures 5.1 and 5.2) are translated into multivariate continuous-time Markovian processes (Markov, 1906), with stochastic variables each representing the number of bacteria in the blood, liver, spleen and MLNs (observed experimental output). The dynamics of such processes, in particular the probability that the system is at a given state at any particular time point, can be summarised by the forward Kolmogorov equation, which describes this probability in terms of the contribution of each of the biological processes included in the mechanistic model (bacterial replication, killing, migration and phenotypic switch). In this case, the state refers to a vector consisting of the number of bacteria in all 4 locations at the time point of interest. Transitions between different states, i.e. any change in the number of bacteria in any of the tissues of interest over time, are driven by dynamical processes quantified by parameters, reflecting the rate of development of unobserved bacterial dynamics in time.

5.2.3.1. Summary statistics for cross-distribution comparison

For a multivariate stochastic system, the overarching forward Kolmogorov equation becomes analytically intractable for multi-compartmental systems with numerous parameters. To avoid the step of solving the probability density function characterising the multivariate bacterial distribution, we use summary statistics to capture the necessary features of interest. Our models only incorporate zero-order dynamics, thus the first- (mean numbers of bacteria in each organ) and second-order (variance-covariance matrix of bacterial numbers in all organs) moments suffice to describe the distribution (Price *et al.*, 2017).

5.2.3.2. Divergence minimisation parameter inference framework

Here, we solve the mathematical models retrospectively using experimentally measured outputs (numbers of bacteria per organ) with the aim to estimate the values of unobserved processes, which correspond to parameters characterising bacterial replication, death, migration and switching to a non-growing phenotype. To achieve this, we use a likelihood-free inference framework, which overrides the problem of intractability. As the multivariate distribution is characterised by a set of first- and second- order moments, it is possible to explore parameter value combinations with the aim of reaching a distribution that approximates the experimentally observed bacterial distribution as closely as possible. This is achieved by minimising a quantity (cost function) that cumulatively expresses the

divergences between each experimentally-observed and predicted pair of moments (Price *et al.*, 2017). In this case we choose the Kullback-Leibler divergence as the cost function to be minimised.

Parameter inference was performed using the freely available SPEEDI.R package (Price and Restif, 2017) and the Nelder & Mead (Nelder and Mead, 1965) optimisation algorithm.

5.2.3.3. Parametric bootstrap for quantification of inference precision and assessment of model goodness-of-fit

Goodness-of-fit is a measure of how well a given model captures the experimental data. Here, we use a parametric bootstrap approach to assess model goodness-of-fit. Using minimum-divergence moments-based inference, we identify the set of parameters that correspond to first- and second- order moments and yield the smallest divergence from the observed moments. This set of moments is referred to as the predicted moments and the corresponding parameter set as the predicted parameters.

Using the predicted parameters, we simulate 500 samples of size emulating the number of observations (10 mice x 8 ITS = 80 or 5 mice x 8 ITS = 40) per time point in the experimental dataset. For each simulated sample, we calculate the corresponding moments and compare each set of moments with the predicted moments to obtain a divergence value. We refer to those moments as simulated moments. If each of the experimental moments fall into their respective distribution of simulated moments, the goodness-of-fit is adequate and the model can be accepted as plausible.

We, then, use the simulated moments to anew infer the range of parameter values, which provide the confidence intervals for our original parameter estimates, and thus serve as a measure of the inference precision.

5.3. Results

5.3.1. Simulation study for the detection of heterogeneity in ITS data

Using infections with mixed inocula comprising ITS it is possible to infer the rates at which inter- and intra-organ dynamics of infection change over time. However, it has not hitherto been shown whether data from ITS-based studies can be used to address questions about heterogeneity in within-host dynamics. Here, we are interested in comparing the performance of a SP model assuming a single bacterial population with uniform dynamics, and that of a DP model assuming two bacterial subpopulations with different dynamics, described in section 5.2.2.

We performed a simulation-based study to investigate whether bacterial populations with uniform replication and death dynamics could be reliably distinguished from bacterial populations composed of both a growing, antibiotic-sensitive subpopulation and a non-growing, antibiotic-recalcitrant subpopulation, by applying mechanistic models to simulated data from synthetic ITS-based "experiments".

We simulated 10 synthetic datasets with different parameter sets from the dual phenotype DP-model. We then applied both the SP- and DP-models to the 10 synthetic datasets to obtain a best parameter estimate, and as per 5.2.2.3, we used a parametric bootstrap approach to generate 500 bootstrapped samples for each of the 10 synthetic datasets under the two models. We finally performed a pairwise comparison of their goodness-of-fit.

We used two tools to evaluate the goodness-of-fit of the SP- and DP- models to the synthetic datasets. The first was the pairwise comparison of the KL divergences; a lower divergence indicates a better fit between the model and the synthetic data. The second tool was the pairwise comparison of the similarity between each synthetic dataset and its bootstrapped samples when the DP- or SP- model was used respectively. To this end, for each of the two models we obtained the mean value and standard deviation of each moment for the 500 bootstrapped samples and calculated the number of standard deviations between each true moment value and the mean value from the bootstrapped samples.

For all 10 synthetic datasets simulated from the DP-model, the DP-model gave consistently lower Kullback-Leibler (KL) divergence values by two orders of magnitude, compared to the SP-model (shown in Table 5.3). When applied to DP-data, the bootstrapped samples

from the DP-model were more similar to the original synthetic data than bootstrapped samples from the SP-model (Figure 5.3).

Overall, this simulation study shows that phenotypic heterogeneity at the level of two subpopulations is detectable from ITS-based data collected at the level of the entire bacterial population.

Synthetic dataset index	KL divergence from DP-model	KL divergence from SP-model
1	0.0031	0.55
2	0.0053	1.38
3	0.0041	0.88
4	0.0013	0.20
5	0.0026	1.17
6	0.0045	0.47
7	0.0017	0.46
8	0.0045	0.45
9	0.0012	0.92
10	0.0046	1.16

Table 5.3: Pairwise comparison of KL-divergence values from DP- and SP- model fitting to synthetic datasets simulated from the DP-model.

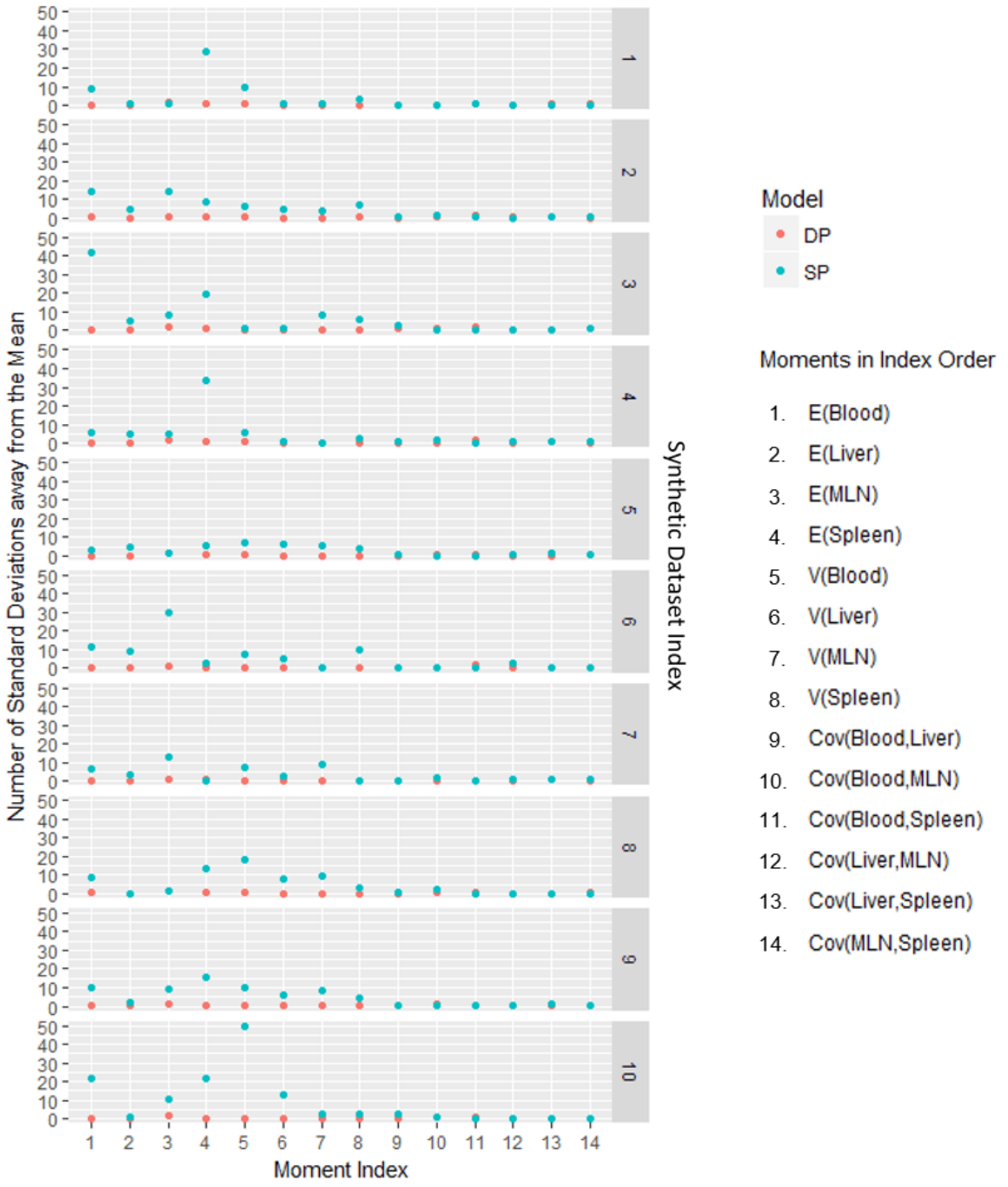


Figure 5.3: Comparison of SP- and DP- model performance on synthetic data simulated from the DP-model.

Each of the horizontal panels (1-10) represents the fit between one synthetic dataset and the SP- and DP-models in blue and red respectively. The blue and red dots represent the number of standard deviations between the moments of the virtual dataset and the mean value of 500 bootstrapped datasets generated from the best parameter set in the SP and DP model respectively. The moment indices in the x-axis refer to the 14 means, variances and covariances used as summary statistics of the ITS distributions in the 4 organs. The number of standard deviations is always near-zero when the DP-model is applied (red), while consistently higher when the SP- model is applied (blue).

5.3.2. Model Selection

First, we identified the best-fitting, most parsimonious model from the family of single-phenotype (SP) models, whose structure is shown in Figure 5.1. Starting from a fully parameterised model (12 parameters), we fitted sequentially simpler models (nested models) to data from ampicillin- and ciprofloxacin treated mice day-by-day for 4 days of total treatment and 2 days post-treatment. For each treatment group, we used the ITS counts per organ at 72 hours post-inoculation as the starting point and the respective ITS counts at 24 hours post-treatment to mark the end of the first day of the antibiotic regimen, and inferred the best parameter estimate set, as detailed in 5.2.5.1. We repeated the same inference process to estimate the parameter values for each time interval separately, totalling four time intervals during antibiotic treatment and 2 during the relapse phase.

We repeated the same process for the family of DP models, whose structure is shown in Figure 5.2. The best-fitting models for the SP and DP family of models are comparatively summarised for each time interval, in the ampicillin- and ciprofloxacin-treated groups. Note that where the DP model offers no improvement in fitting the experimental data, the best DP model converges to the best-fitting SP model (Tables 5.4 and 5.5 respectively).

Model	Treatment Group	Time Interval (days after start of treatment)	Parameters Estimated	Number of Parameters Estimated	KL divergence
SP	Ampicillin	1	k1,k2,k3,k4,r2,r3,r4,m4.2	8	0.66
DP	Ampicillin	1	k1,k2,k3,k4,r2,r3,r4,m4.2	8	0.63
SP	Ampicillin	2	k1,k2,k3,k4,r2,r3, r4	7	0.98
DP	Ampicillin	2	k1,k2,k3,k4,r2,r3, r4	7	0.93
SP	Ampicillin	3	k1,k2,k3,k4,r2r3,r4	7	1.60
DP	Ampicillin	3	k1,k2,k3,k4,r2,r3,r4,m4.8,m8.6	10	0.56
SP	Ampicillin	4	k1,k2,k3,k4,r3,r4, m1.4	7	0.50
DP	Ampicillin	4	k1,k2,k3,k4,r1,r2,r3,r4,m3.7,m4.8	10	0.49
SP	Ampicillin	5	k2,k3,k4,r1,r2,r3,r4	8	0.58
DP	Ampicillin	5	k2,k6,k7,k8,r1,r2,r6,r7,r8	9	0.12
SP	Ampicillin	6	k1,k2,k3,k4,r1,r2,r3,r4	8	0.73
DP	Ampicillin	6	k1,k2,k6,k7,k8,r1,r2,r6,r7,r8	10	0.39

Table 5.4: Comparative summary of best-fitting SP- and DP- models in the ampicillin-treated mice

Model	Treatment Group	Time Interval (hours after start of treatment)	Parameters Estimated	Number of Parameters Estimated	KL divergence
SP	Ciprofloxacin	0-24	k1,k2,k3,k4,r1,r2,r3,r4	8	0.06
DP	Ciprofloxacin	0-24	k2,k4, r2,r4,m3.7,m1.2	6	0.40
SP	Ciprofloxacin	24-48	k1,k2,k3,k4,r1,r2,r3,r4,m2.4	8	2.91
DP	Ciprofloxacin	24-48	k1,k2,k4,r2,r4,m2.4,m3.7,m4.8, m8.6	8	3.40
SP	Ciprofloxacin	48-72	k1,k2,k3,k4,r1,r2,r3,r4,m2.4	9	0.25
DP	Ciprofloxacin	48-72	k1,k2,k3,k4,r1,r2,r3,r4,m1.4,m4.8, m2.6,m8.4,m6.2,m7.3	14	0.33
SP	Ciprofloxacin	72-96	k2,k3,k4,r1,r2,r3,r4	7	0.14
DP	Ciprofloxacin	72-96	k2,k3,k4, r1,r2,r3,r4,m4.8,m8.4, m6.2	10	0.19
SP	Ciprofloxacin	96-126	k2,k3,k4,r1,r2,r3,r4	7	0.54
DP	Ciprofloxacin	96-126	k2,k3,k4,k5,k7,k8,r2,r3,r4,r5,r7,r8, m6.8	13	0.85
SP	Ciprofloxacin	126-140	k1,k2,k3,k4,r1,r2,r3,r4,m2.4	9	0.38
DP	Ciprofloxacin	126-140	k1,k2,k3,k4,r1,r2,r3,r4,m2.4	9	0.38

Table 5.5: Comparative summary of best-fitting SP- and DP- models in the ciprofloxacin-treated mice

Having identified the best-fitting model from each model family for each of the 6 time intervals considered in each of the 2 treatment groups, we then assessed whether the divergence of the best predicted model from the experimental data can be recapitulated by bootstrapped datasets simulated from the best parameter set corresponding to the best fitting models. Given that the DP model family is more complex than the SP model family, we chose the best-fitting DP model as superior to the best-fitting SP model only if bootstrapped data from the former were consistently more similar to the experimentally observed data than synthetic data from the SP model.

To quantify the similarity between the experimentally observed and bootstrapped data produced from either the best-fitting SP or the best-fitting DP models, we calculated the first- and second-order moments for each of the 500 bootstrapped datasets produced from each of the best-fitting models per time interval per treatment group, followed by the mean and standard deviation for each of those moments. As illustrated in Figure 5.4, we then calculated in standard deviations the distance between the moments from the experimentally observed data and the mean values for the moments of the bootstrapped samples, similarly to 5.3.1.

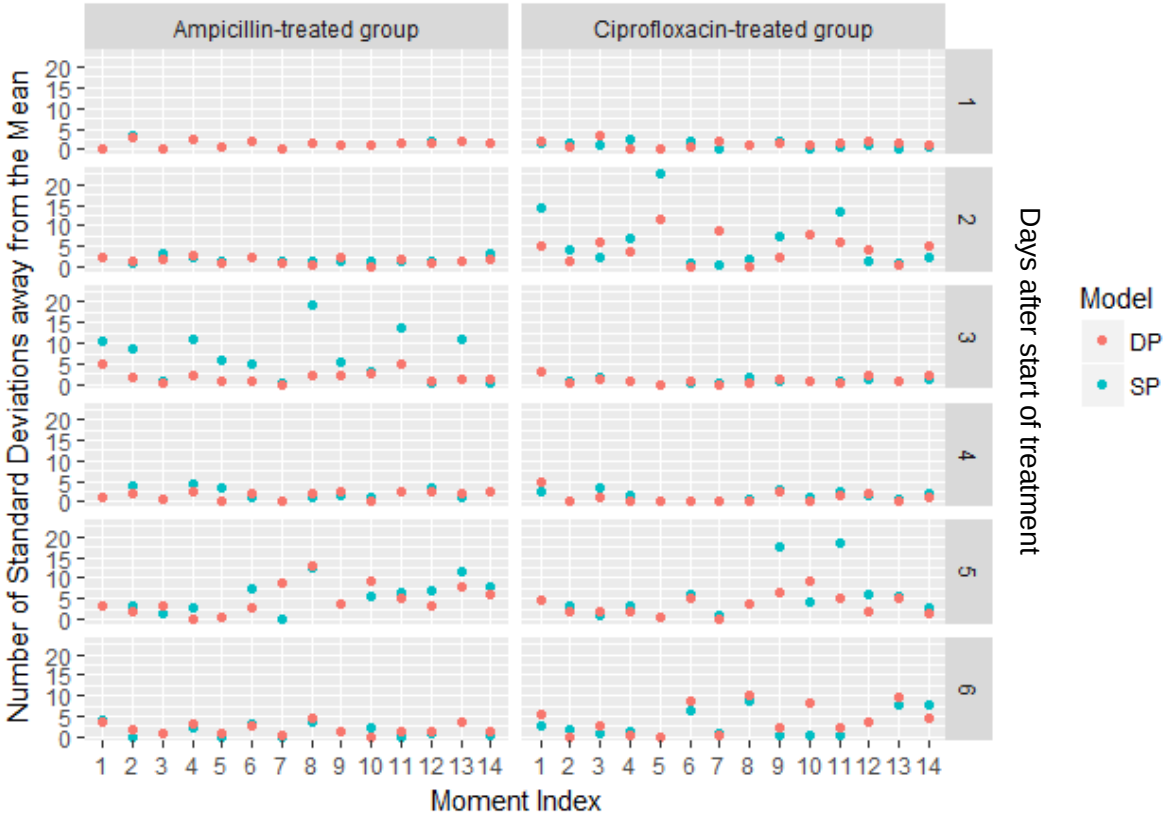


Figure 5.4: Comparison of SP and DP model performance assessing the relationship between model-based simulations and observed experimental values

The blue and red dots represent the number of standard deviations between the experimentally observed moments and the mean value of 500 bootstrapped datasets generated from the best parameter set in the SP and DP model respectively, for each time point. The indices in the x-axis refer to the 14 means, variances and covariances used as summary statistics of the ITS distributions in the 4 organs. The enumeration of the horizontal panels (1-6) refers to the number of days after the start of antibiotic treatment. In ampicillin-treated mice, improvement in the model fit with the DP model is shown by an overall reduction of the number of standard deviations between the experimentally observed moments and the mean value of 500 parametric datasets, especially during days 2, 3 and 5. In ciprofloxacin-treated mice, there is no consistent improvement when the DP model is applied.

5.3.3. SP and DP model comparison in ampicillin-treated mice

We applied the SP model, described in section 5.2.2., to the dataset from the ampicillin-treated group. Upon obtaining the best parameter estimates for each time interval, we simulated 500 bootstrapped datasets to obtain a distribution of values for each of the moments. We expected that in a well-fitting model, the experimentally observed moments would fall within the distribution of bootstrapped moments. Figure 5.5 summarises the results of the fit of the DP- and SP- model to the experimental data, showing that the DP-model improves the fit of the model to the experimental data, on days 3-5 after the start of treatment.

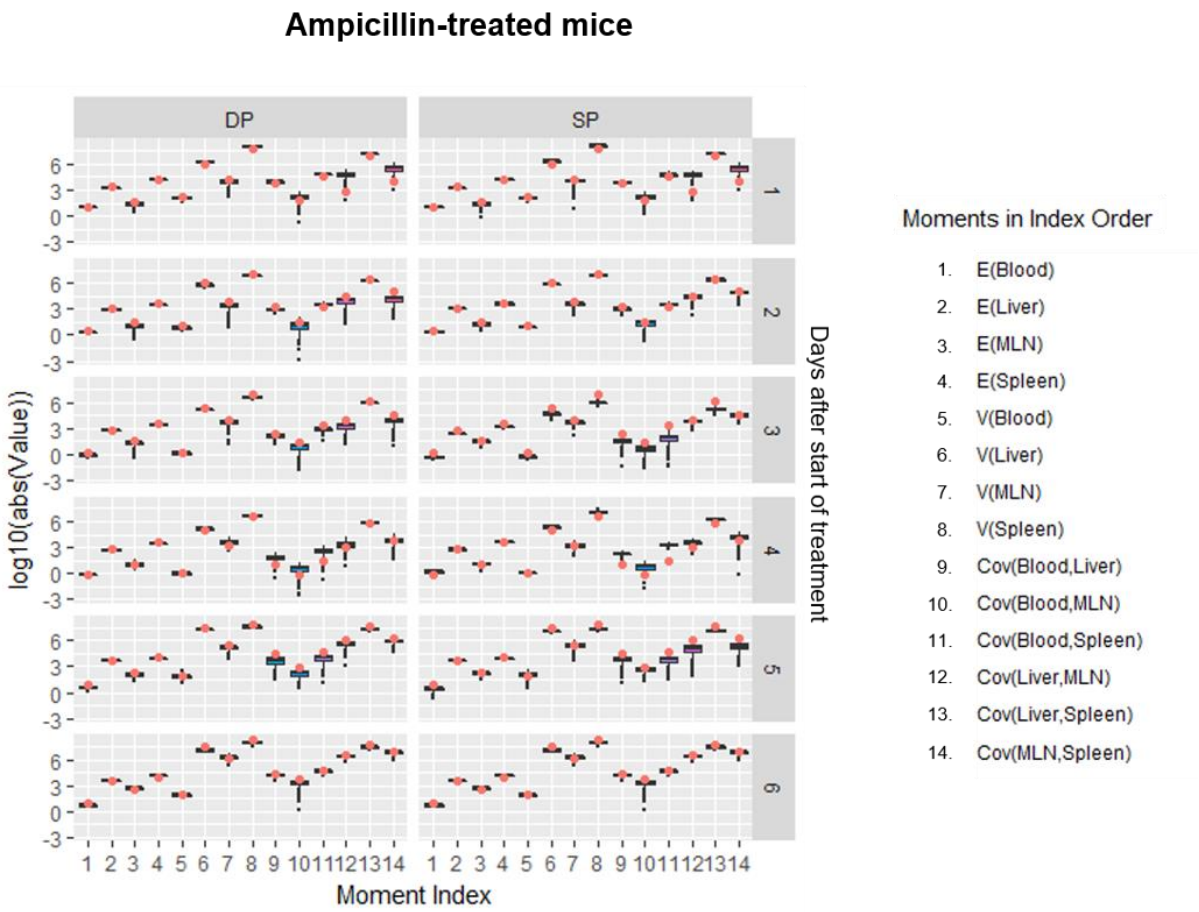


Figure 5.5: Comparison between the fit of the SP- and DP- models to the experimental data.

Each of the boxplots captures the distribution of 500 bootstrapped values for each of the 14 first- and second-order moments, while each red dot represents the value of the same moment from the experimentally observed data. Improvements can be seen in moments with indices 8, 9, 10, 11 and 14 on day 3, moments with indices 1, 6, 8, 9, 10, 11 and 13 on day 4, and moments with indices 6, 8, 12 and 14 on day 5. The improvement mostly concerns the variances (indices 4-8) and covariances (indices 8-14).

5.3.4. SP model-fit in ciprofloxacin-treated mice

We then applied the SP-model, described in section 5.2.2., to the dataset from the ciprofloxacin-treated group and obtained a distribution of values for each of the moments, as in 5.3.2. Here, we show that throughout the treatment and relapse phase, the experimentally observed moments can be recapitulated by simulations from the SP-model given the best set of parameter estimates (Figure 5.6).

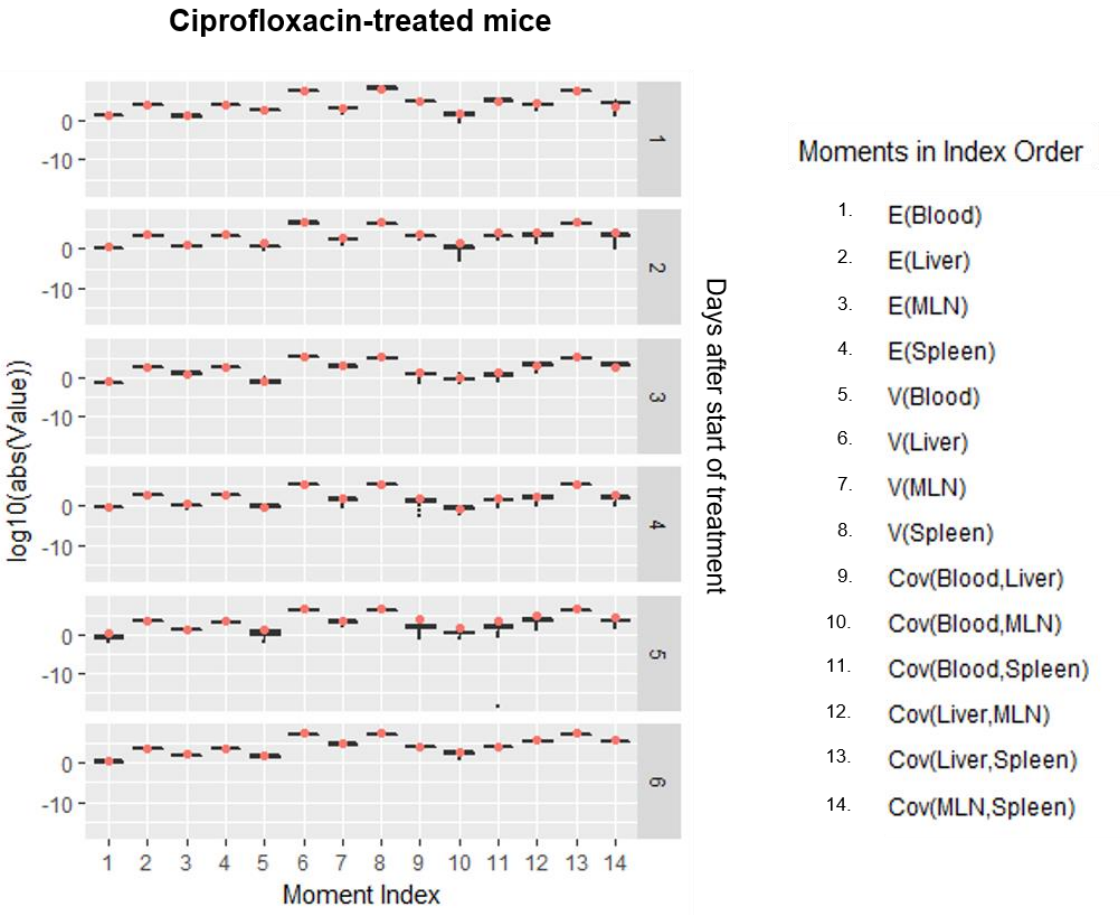


Figure 5.6: Experimental moments in relation to moments from synthetic data using the best-fitting SP model.

5.3.5. Intra-organ dynamics during the treatment phase

Our models predict that ampicillin and ciprofloxacin exert different effects on intra-organ bacterial dynamics in terms of their bactericidal and bacteriostatic effects (Figure 5.7). While in ciprofloxacin-treated mice, bacterial replication occurs throughout the duration of treatment in all four organs, ampicillin suppresses bacterial replication, particularly in the blood and in the late phase of treatment in the spleen and MLN. The occasionally wide confidence intervals in our estimates are explained by the correlation between replication and killing rates, as shown in by the narrow ellipses in Figure 5.8.

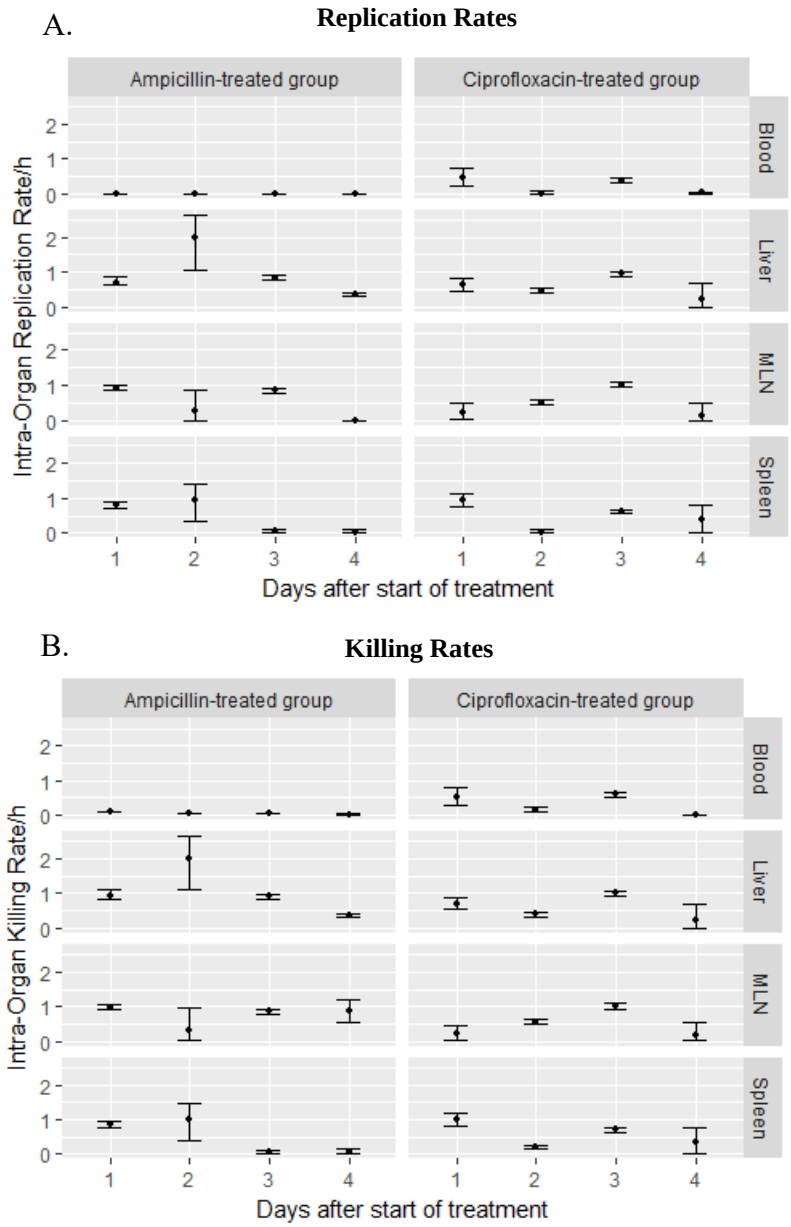


Figure 5.7: 95% confidence intervals of intra-organ replication and killing rates estimated from 500 parametric bootstrapped samples from the best parameter estimates.

Panel A comparatively shows the replication rates for ampicillin-treated and ciprofloxacin-treated groups, while panel B shows the killing rates for the two groups. The dots represent the mean values across 500 parametric bootstrapped samples from the best parameter estimates, while the bars represent the lower and upper boundaries of the 95% confidence intervals. The cidal activity in ciprofloxacin-treated mice culminates after 3 days of treatment, while the static effects of ampicillin peak after 4 days of treatment.

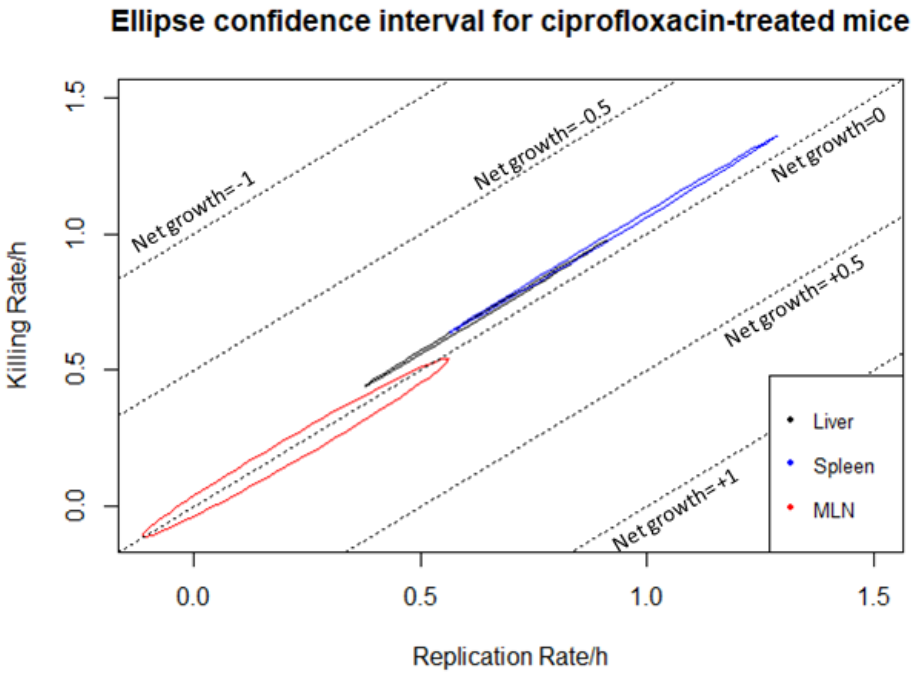


Figure 5.8: Bivariate 95% confidence regions of intra-organ replication and killing rates (500 bootstrapped samples) for ciprofloxacin-treated mice 1 day after treatment.

The black, blue and red ellipses illustrate the bivariate 95% confidence regions for replication and killing rate in the liver, spleen and MLN respectively during day 3 after the start of ciprofloxacin treatment. The large length of the major axis of the ellipse represents the high correlation between replication and killing parameter estimates, which accounts for the large confidence intervals seen in Figure 5.7. Diagonal grid lines indicate net growth rates (replication rate-killing rate). In the liver and spleen, cidal effects are stronger leading to a negative net growth rate, while net growth in the MLN is close to 0.

5.3.6. Non-growing subpopulations recalcitrant to antibiotic therapy in ampicillin-treated mice

In ampicillin-treated mice, the best-fitting DP-model predicts the emergence of a non-growing, antibiotic-recalcitrant bacterial subpopulation. Our inference shows that this bacterial subpopulation emerges 48 hours after treatment with ampicillin at varying proportions in different organs and over time. By 72 hours of treatment, non-growing antibiotic-recalcitrant bacteria originating from the spleen have migrated to the liver. By 96 hours, the entirety of the bacterial population in the MLN and spleen have switched to the non-growing, antibiotic-insensitive phenotype, as summarised in Figure 5.9.

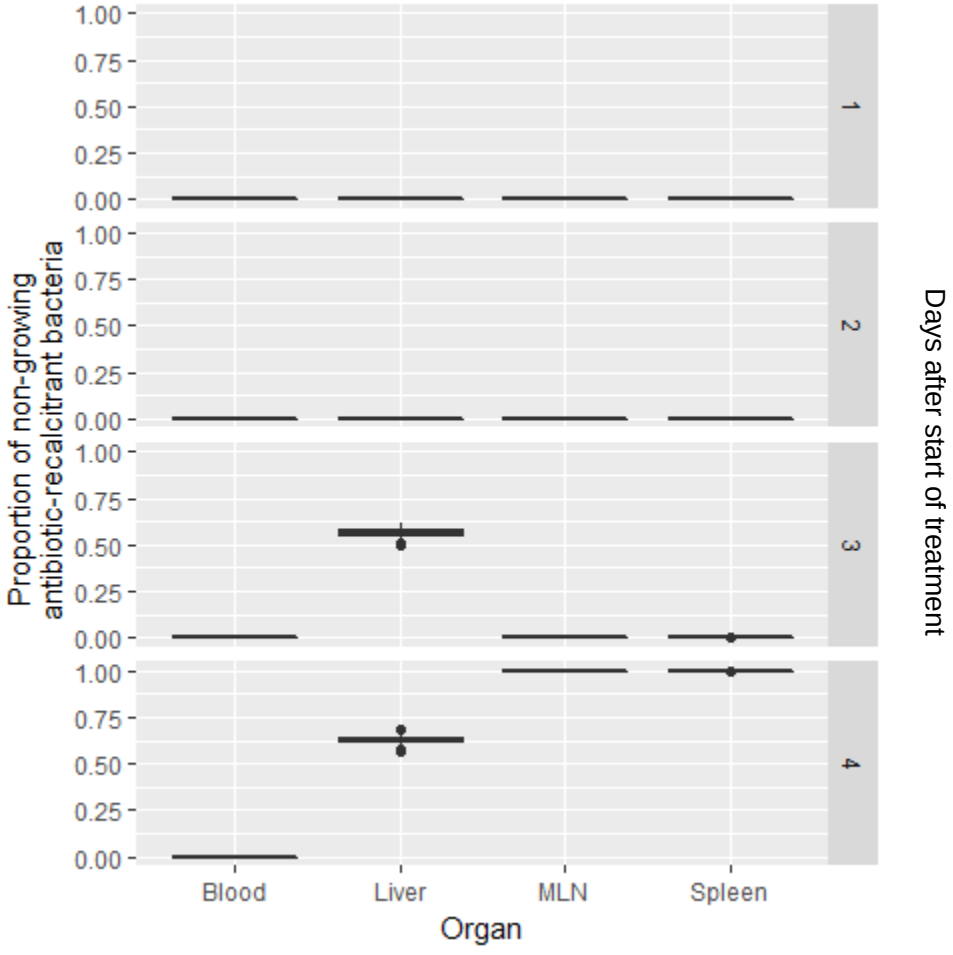


Figure 5.9: Proportion of persister bacteria in the blood, liver, MLN and spleen during ampicillin-treatment.

The confidence intervals for the proportion of persister bacteria are expressed as a fraction of total bacterial number in the liver, MLN and spleen (500 bootstrapped samples).

5.3.7. Dynamics in the relapse phase

Following withdrawal of the 4-days of antibiotic treatment, bacterial infection relapses in both ampicillin- and ciprofloxacin-treated groups, with positive net growth rates in the first 30 hours post-withdrawal. In the ampicillin-treated group, the progeny of the growing subpopulation has a different net growth to that of the non-growing, antibiotic-recalcitrant one in the liver (Liver (P)), MLN (MLN (P)) and spleen (Spleen (P)), shown in figure 5.10.

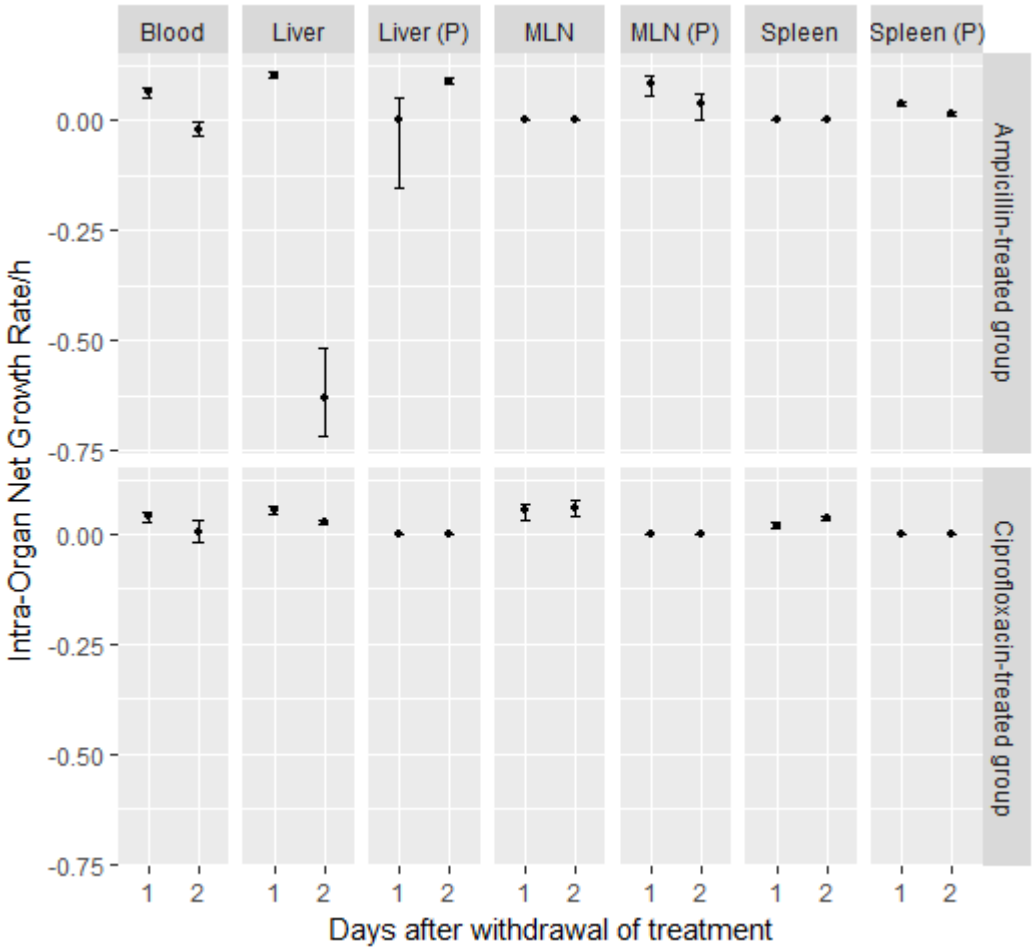


Figure 5.10: 95% confidence intervals of net growth rates (replication - killing rate) during relapse, estimated from parametric bootstrapped samples based on the best set of parameter estimates.

In the ciprofloxacin-treated group with the predicted single bacterial phenotype, the relapse phase is characterised by positive net growth across all organs, with small fluctuations during the 2 days post ciprofloxacin withdrawal. In the ampicillin-treated group, the comparison between the net growth rates for the progenitors of the previously non-growing subpopulation with those of their previously growing counterparts in the liver shows that the former are consistently higher than the latter. Following bacterial relapse, all organs are repopulated by progenitors of the previously non-growing, antibiotic-recalcitrant subpopulation.

To understand what drives the extinction of the previously growing antibiotic-sensitive population in the liver of ampicillin-treated mice, we compare the killing and replication rates between the progeny of the previously growing and the persister population (Figure 5.11).

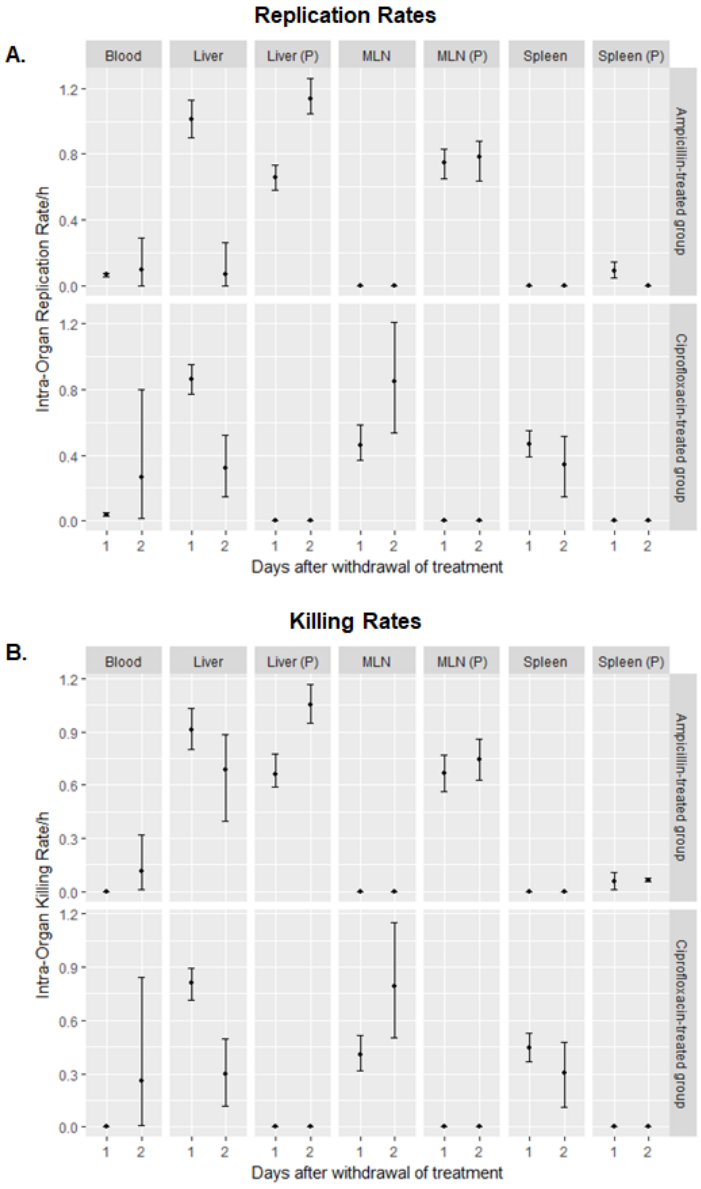


Figure 5.11: 95% confidence intervals of intra-organ replication and killing rates estimated from 500 parametric bootstrapped samples from the best parameter estimates during the relapse phase.

Panel A comparatively shows the replication rates for ampicillin-treated and ciprofloxacin-treated groups, while panel B shows the killing rates for the two groups. The dots represent the mean values across 500 parametric bootstrapped samples from the best parameter estimates, while the bars represent the lower and upper boundaries of the 95% confidence intervals. In the ampicillin-treated mice, progeny of the non-growing antibiotic-recalcitrant bacterial reservoir in the liver (Liver (P)) replicate at a higher rate on average compared to the progenitors of their previously growing counterparts (Liver).

5.4. Discussion

In the present data-based modelling study, we set out to quantify the effects of ciprofloxacin and ampicillin on the *in vivo* dynamics of *S.* Typhimurium in the murine host. Complementing previous studies, this is a mathematical modelling study based on data collected at the level of the bacterial population, using a virulent SL1344 *Salmonella* strain in C57BL/6 immunocompetent mice that have not been pre-treated with antibiotics. We used data from mouse infections with two different classes of antibiotics, one beta-lactam and one fluoroquinolone to directly compare their *in vivo* effects on the bacterial dynamics in an intravenous infection of wild-type mice, closely mimicking the systemic phase of non-typhoidal salmonellosis (Rossi *et al*., 2017).

The inferential capacity of our mechanistic models is dependent on quantitative signals present in the experimental data collected by Rossi *et al.* (Rossi *et al*., 2017) i.e. changes in the mean number of ITS copies per organ during time, their variance and their joint variability across all organs (covariance). Mathematical models can make use of the direction, magnitude and collective effect of these changes to maximise the insight provided by experimental data, especially if the changes in the mean numbers, which constitute the primary tool for qualitative inference, are not strikingly dissimilar. Our simulation study shows that when a single-phenotype model is applied to data from a dual-phenotype system, there is consistent misestimation of the variances across all 10 synthetic datasets. In our experimental data, the ill fit of the SP-model during days 3 and 4 of treatment mainly concerns variances and covariances. It, thus, appears that the quantitative signal allowing us to carry out this inference lies in the second-order moments (variances, covariances) of the ITS distribution in the organs.

As a proof-of-concept, we first conducted a simulation-based study to show that mechanistic models applied to ITS-based data can distinguish between bacterial populations with uniform and heterogeneous dynamics. Then, we fit a both a SP- and a DP- model to data from ampicillin- and ciprofloxacin-treated mice. Our best-fitting models suggest that a SP population can explain the dynamics in ciprofloxacin-treated mice, while the dynamics in ampicillin-treated mice are further complicated by bacterial phenotypic heterogeneity, best captured by the DP model with a non-growing, antibiotic-recalcitrant bacterial subpopulation. Inferences made regarding the rates at which these dynamics develop over

time indicate that while ciprofloxacin acts predominantly via cidal mechanisms, ampicillin has mixed bacteriostatic and bactericidal effects (Landman, 1997) At the end of ampicillin treatment, surviving bacteria in the MLN and spleen have adopted a non-replicating, antibiotic-recalcitrant phenotype; in contrast, the vast majority of surviving bacteria in the ciprofloxacin-treated group are of the growing, antibiotic-sensitive phenotype. Following antibiotic withdrawal, bacteria relapse in both the ampicillin- and ciprofloxacin-treated groups of mice. In ciprofloxacin-treated mice, growth rate in the blood, liver and spleen is higher on day 1 post- antibiotic withdrawal and, subsequently decreases. In ampicillin-treated mice, the liver, spleen and MLN are repopulated by progeny of the non-growing antibiotic-recalcitrant bacteria.

These results can be discussed in the context of previous experimental studies using alternative observational techniques, including non-phenotypic marker-based techniques e.g. Kaiser *et al*., 2014; Diard *et al*., 2014) fluorescence dilution (Helaine *et al*., 2014) and peak-to-trough ratio (Haugan, Løbner-Olesen and Frimodt-Møller, 2018). An *in vitro* study by Ghooi and Thatte (Ghooi and Thatte, 1995) showed that beta-lactam antibiotics, such as ampicillin, act as both bacteriostatic and bactericidal agents, due to their concomitant action on bacterial cell wall synthesis inhibition preventing bacterial replication, and their action on autolysins inducing bacterial death respectively. McCall *et al.* (McCall *et al.,* 2019*)* introduced the phenomenon of antibiotic-induced stasis to explain the emergence of a slow- or non-growing bacterial population as a result of exposure to antibiotics with bacteriostatic effects and proceeded to show that this population cannot be killed by the same antibiotic class. The inferences made by our best-fitting DP model based on data from ampicillin-treated mice aligns with these documented *in vitro* effects of ampicillin on bacterial growth, as we estimated both strong bacteriostatic effects in the blood and the organs after 48 hours of treatment, and the emergence of a non-growing, antibiotic-recalcitrant subpopulation in organs where persisters have previously been reported *in vivo* (Lewis, 2010).

In vivo studies using quinolones such as enrofloxacin and ciprofloxacin have highlighted the emergence of non-growing bacteria that survive treatment in the mesenteric and caecal lymph nodes, often termed persister cells (Claudi *et al*., 2014; Kaiser *et al*., 2014; Helaine *et al*., 2014; Griffin *et al*., 2011). In these studies, fluoroquinolones were used very soon after inoculation, and at high doses (8mg/mouse enrofloxacin oral in Kaiser et al. (Kaiser *et al*., 2014); 2.4 mg/mouse ciprofloxacin intraperitoneal in Helaine *et al.* (Helaine *et al.,* 2014).

Fluoroquinolones are known to retain their cidal activity against non-replicating bacteria *in vitro*, hence additional factors such as the high antibiotic concentration or low oxygen tension *in vivo* could be responsible for the emergence of persister bacteria in the gut-associated lymphatic tissues (Lewin, Morrissey and Smith, 1991).

Helaine *et al.* (Helaine *et al.*, 2014) used a fluorescence dilution (FD) technique to explore the replication history of individual bacteria during exposure to early, high-dose enrofloxacin in different tissues. They identified persister bacterial populations in the lymph nodes and spleen. Our modelling study indicates that the spleen and MLN are indeed the main sites of interest regarding the emergence of persisters, when mice were treated with ampicillin. However, our model did not predict this persisting subpopulation in ciprofloxacin-treated mice. This discrepancy could either be due to the low levels of persisters in this group, which would not affect the ITS distributions significantly enough to be detected by our model, or due to the differences in the experimental design of the studies, especially the timing and dose of the fluoroquinolone therapy, which is known to affect its pharmacokinetic properties (Lewin, Morrissey and Smith, 1991). Our findings suggest that at clinically relevant doses, non-growing "persisters" may not be a universal sequela of antibiotic treatment, but may be antibiotic-specific, as our SP-model was still adequate to capture the dynamics in the ciprofloxacin-treated group. Alternatively, the vast majority of persister bacteria following quinolone exposure may originate from growing instead of non-replicating bacteria (Claudi *et al.*, 2014). Finally, an important determinant in the apparent lack of consensus in this field is the significant inherent stochasticity in extinction time and dynamics *in vivo* during antibiotic treatment even for large bacterial populations, compounded by host variability (Coates *et al.*, 2018).

As noted, the lymph nodes and spleen have been previously identified as organs that harbour non-replicating bacteria recalcitrant to antibiotic activity of beta-lactam antibiotics. Our best-fitting DP model for the ampicillin-treated group further suggests that while this group of persister bacteria originate from bacterial populations originally residing in the spleen and MLN, they have migratory potential. In particular, these non-growing bacteria migrate from the spleen to the liver later during the treatment, and interestingly, once in the liver they convert back to the original phenotype, regaining their replicative capacity, but also their sensitivity to antibiotic killing. While the trigger of the switch is unknown at this stage, one candidate mechanism could be the change in bacterial microenvironment due to inter-tissue

migration, particularly the change in occupied host cell type. In a recent *in vivo* study of splenocytes infected with *S.* Typhimurium at the single cell level, Kanvatirth *et al.* (Kanvatirth *et al.*, 2020) showed that different host cell types affect not only how easily *Salmonellae* can replicate or become a target for killing intracellularly, but also how readily they enter a low metabolic, persister state. In contrast to the intimate relationship between the liver and the spleen, recalcitrant bacteria arising in the MLN do not appear to migrate either in the liver or spleen, similar to their growing, antibiotic-sensitive counterparts in that regard. As a result, the MLN stands in isolation, showing distinct dynamics, while the liver and spleen form a network supporting continuous inter-organ migration.

During the relapse phase, following withdrawal of antibiotic therapy, bacteria resume growth and repopulate the liver, spleen and MLN. In the livers of ampicillin-treated mice where the antibiotic-sensitive and antibiotic-recalcitrant subpopulations co-exit, the former is killed at higher rates despite replicating at lower rates than the latter, collectively leading to its extinction. As a result, all three tissues are repopulated by progeny of the antibiotic-recalcitrant subpopulation.

Nevertheless, inferences made by our model are subject to three levels of limitations, which we describe here. Limitations at the level of the experimental data include the absence of markers of bacterial metabolic activity and uniquely identifiable persister subpopulations. In terms of our modelling approach, we assumed the simplest case of linear rates describing the underlying dynamics of the experimental system; although simplifying, our assumption is justified by dividing the infection timeline into shorter time intervals and making independent inferences for each, and the fact that we only estimated average rates. We developed and applied a single- and a dual-phenotype model but did not increase the model complexity further to include more than two bacterial phenotypes. Our decision was based on the law of parsimony, whereby the simplest possible model structure that adequately captures the observed data is preferred; this reduces the risk of overfitting and model unidentifiability. Additionally, we pooled the ITS distributions from multiple mice per treatment group per time point, effectively assuming homogeneous host immune responses. Finally, at the level of interpreting our results, we have attributed cidal activity to antibiotic effects, ignoring any baseline immune responses. For the purposes of comparing the effects of the two antibiotics, and if the assumption that the average baseline immune response

between the two groups is similar, then the differences in the dynamics of two groups are still largely rightly attributable to the action of the antibiotic in question.

In summary, in this paper we present an integrated approach between experimental data and mechanistic mathematical models in a case of immunologically intact, previously untreated mice infected with *S.* Typhimurium, treated with either ampicillin or ciprofloxacin, and finally allowed to relapse. Our study showcases the value of mechanistic mathematical models in maximising biological insight from experimental data, where previous qualitative inference only provides limited information. The additional benefit of our data-driven modelling approach is making inferences about processes that cannot be or remain challenging to observe experimentally in a living organism. External validation of our inferential insights would require *in vivo* experimental setups that involve groups independently treated with different antibiotics at clinically relevant regimens, and a reliable marker to discern non-growing, antibiotic-recalcitrant bacteria at all time points. Ideally, fine-scale data on drug penetration in different tissues, local drug concentration gradients and host cell uptake would be needed to determine the exact microenvironment to which bacteria are exposed. Serial sampling of the bacterial load from the same host would eliminate any variation attributable to within-host heterogeneity in immune responses.

6.

An experimental design tool to optimise inference precision in data-driven mathematical models of bacterial infections *in vivo*

Summary Points:

- With early integration of mathematical modelling in the experimental design process, it is possible to identify experimental designs which maximise the information content per experimental resource unit.
- Optimal experimental designs aid the decision-making process around optimal resource allocation.
- Precise inference through integrating experimental data with mathematical models enables the clear-cut comparison between competing biological hypotheses and the generation of meaningful predictions.
- Tools to quantify the information of alternative experimental designs have been lacking in microbiology due to the long-standing lack of computationally efficient inference methods.
- Inspired by the canonical D-optimality criterion, I choose the determinant of the parameter estimate variance-covariance matrix as a metric for global variance in the inferred parameters in a likelihood-free inference framework applied to ITS data.
- Prediction of high-precision experimental designs is more consistent in biological systems with low variance due to low bacterial replication and killing rates.
- A higher number of biological replicates yields higher-precision inference.
- A lower number of ITS per animal host, leading to higher number of copies per ITS yields higher-precision inference.
- Given a constraint on the total number of animal hosts used, a lower number of biological replicates with a low number of ITS per host leads to higher-precision inference than a higher number of biological replicates with a high number of ITS per host.

Declaration of Originality

This chapter has been reproduced verbatim from a published research paper (Vlazaki. Price and Restif, 2020), for which I certify that I am the sole contributor in terms of study design, execution, and manuscript writing. I also certify that I have included the relevant citations, where appropriate.

6.0. Abstract

The management of bacterial diseases calls for a detailed knowledge about the dynamic changes in host-bacteria interactions. Biological insights are gained by integrating experimental data with mechanistic mathematical models to infer experimentally unobservable quantities. This inter-disciplinary field would benefit from experiments with maximal information content yielding high-precision inference. Here, we present a computationally efficient tool for optimising experimental design in terms of parameter inference in studies using isogenic tagged strains. We study the effect of three experimental design factors: number of biological replicates, sampling timepoint selection and number of copies per tagged strain. We conduct a simulation study to establish the relationship between our optimality criterion and the size of parameter estimate confidence intervals, and showcase its application in a range of biological scenarios reflecting different dynamics patterns observed in experimental infections. We show that in low-variance systems with low killing and replication rates, predicting high-precision experimental designs is consistently achieved; higher replicate sizes and strategic timepoint selection yield more precise estimates. Finally, we address the question of resource allocation under constraints; given a fixed number of host animals and a constraint on total inoculum size per host, infections with fewer strains at higher copies per strain lead to higher-precision inference.

Relevant published output:

- Vlazaki M, Price DJ, Restif O. 2020 An experimental design tool to optimise inference precision in data-driven mathematical models of bacterial infections in vivo. Journal of The Royal Society Interface, 17(173), 20200717. https://doi.org/10.1098/rsif.2020.0717

6.1. Introduction

Bacterial infections remain leading causes of mortality and morbidity worldwide, accounting for > 50% of annual deaths in vulnerable demographic groups such as the elderly, young children and immunocompromised individuals, and disproportionately afflicting resource-constrained regions (World Health Organisation, 2019). At the same time, the rapid emergence of antibiotic resistance (Hofer, 2018) and tolerance (Brauner *et al*., 2016) are swiftly incapacitating an increasing number of clinically licenced antibiotic agents. Research and development of both novel antibiotics and vaccines is slow, leaving us exposed to untreatable, chronic and recrudescent bacterial infections.

The establishment, progression and outcome of a bacterial infection in a host is shaped by the dynamics of bacterial transmission, colonisation, within-host replication, death, inter-tissue migration, and the interaction of bacteria with the immune system or administered therapeutic agents. Deciphering each of these factors individually can accelerate the transition from empirical to targeted engineered approaches in bacterial treatment and prophylaxis (Coward *et al*., 2014). To aid the characterisation of these dynamics, we often combine high-resolution experimental techniques with inferential mathematical models that fill in the gaps with regards to processes that remain experimentally unobservable (Restif, Thakar and Harvill, 2009).

In the field of bacterial dynamics, isogenic-tagged strains (ITS) constitute one of the most widely used experimental techniques to obtain *in vivo* data about the dynamic behaviour of bacteria in the host (Abel *et al*., 2015). ITS-based data have been historically paired with mathematical models to maximise the biological insight they can offer (Vlazaki, Huber and Restif, 2019; Rossi *et al*., 2020). By complementing experimental data with mechanistic mathematical models, we go beyond the mere characterisation of statistical differences and correlations in the data, to formulation of biological hypotheses into mathematical relations and inference-making by fitting the models to the experimental observations. Representative examples of integrating such models with ITS-based data include studies identifying bottlenecks during the colonisation and infection process (Abel *et al*., 2015; Kaiser *et al*., 2014; Kaiser *et al*., 2013; Grant *et al*., 2008), mapping out the host immune responses underpinning defence at different stages of the infection (Kaiser *et al*., 2013; Grant *et al*., 2008), deciphering the effects of host adaptation in subsequent infections (Dybowski *et al*.,

2015; Dybowski *et al*., 2017), comparing the effects of different vaccine formulations (Coward *et al*., 2014) and antibiotic therapies (Vlazaki *et al*., 2020; Kaiser *et al*., 2014) on infection control, characterising the within-host dynamics of co-infections (Abel *et al*., 2015) and understanding the directionality of bacterial migration and *de novo* tissue colonisation *in vivo* (Kaiser *et al*., 2013; Kaiser *et al*., 2014).

The aim of statistical inference is to estimate biological parameters as accurately and precisely as possible. Precision is typically reported in the form of confidence intervals (Vlazaki *et al*., 2020; Dybowski *et al.*, 2015; Kaiser *et al*., 2014; Coward *et al*., 2014), which are instrumental to assess the evidence for or against hypotheses such as whether bacteriostatic or bactericidal effects exerted by the host's immune system or by a therapeutic agent dominate different stages in the infection process. Quantifying parameter estimate uncertainty is also important when we use a parameterised model for predictive purposes, as it allows us to predict the range of the potential future trajectories of the infection. To enable predictions and comparisons to be as clear-cut and distinguishable as possible, it is desirable to obtain parameter estimates with the narrowest possible confidence intervals.

Although mathematical models have been increasingly used in the study of *in vivo* bacterial population dynamics, their application has until recently remained limited to inference in simple biological systems for short time intervals post-inoculation, typically in the range of 0-12 hours, largely due to computational inefficiencies (Dybowski *et al.*, 2015; Coward *et al.*, 2014; Kaiser *et al.*, 2014; Grant *et al.*, 2008). A recently developed statistical inference tool custom-made for making inferences based on experimental data from studies using ITS (Price *et al*., 2017) has enabled inference in more complex biological systems, with larger inocula, and for the entirety of the infection timeline (Vlazaki *et al*., 2020). This moments-based, divergence minimisation approach represents a computationally efficient inference tool suitable for complex biological models, where traditional maximum likelihood approaches become too computationally expensive.

To this date, there are no published studies providing executable algorithms for extending the application of data-based mechanistic models to experimental design. The availability of a computationally efficient inference method now paves the way for an early iterative integration of mathematical modelling with data collection at the experimental design stage; this can optimise the inferential capacity of models and maximise the biological insights

they generate (Figure 6.1). Model-driven experimental design can also address the long-standing challenge of resource allocation, especially with regards to experimental animal use. In particular, the "Three Rs" tenet — Replacement, Reduction and Refinement (Russell and Burch, 1959) — calls for "Reduction Alternatives" encouraging strategies to reduce the number of animals needed to provide answers to scientific questions; maximising the information obtained per animal has been recognised as one avenue to achieve this (Fenwick, Griffin and Gauthier, 2009). In this paper, we show that with experimental design tools, it is possible to evaluate the information content of alternative designs and identify those with the maximum information content per animal host.

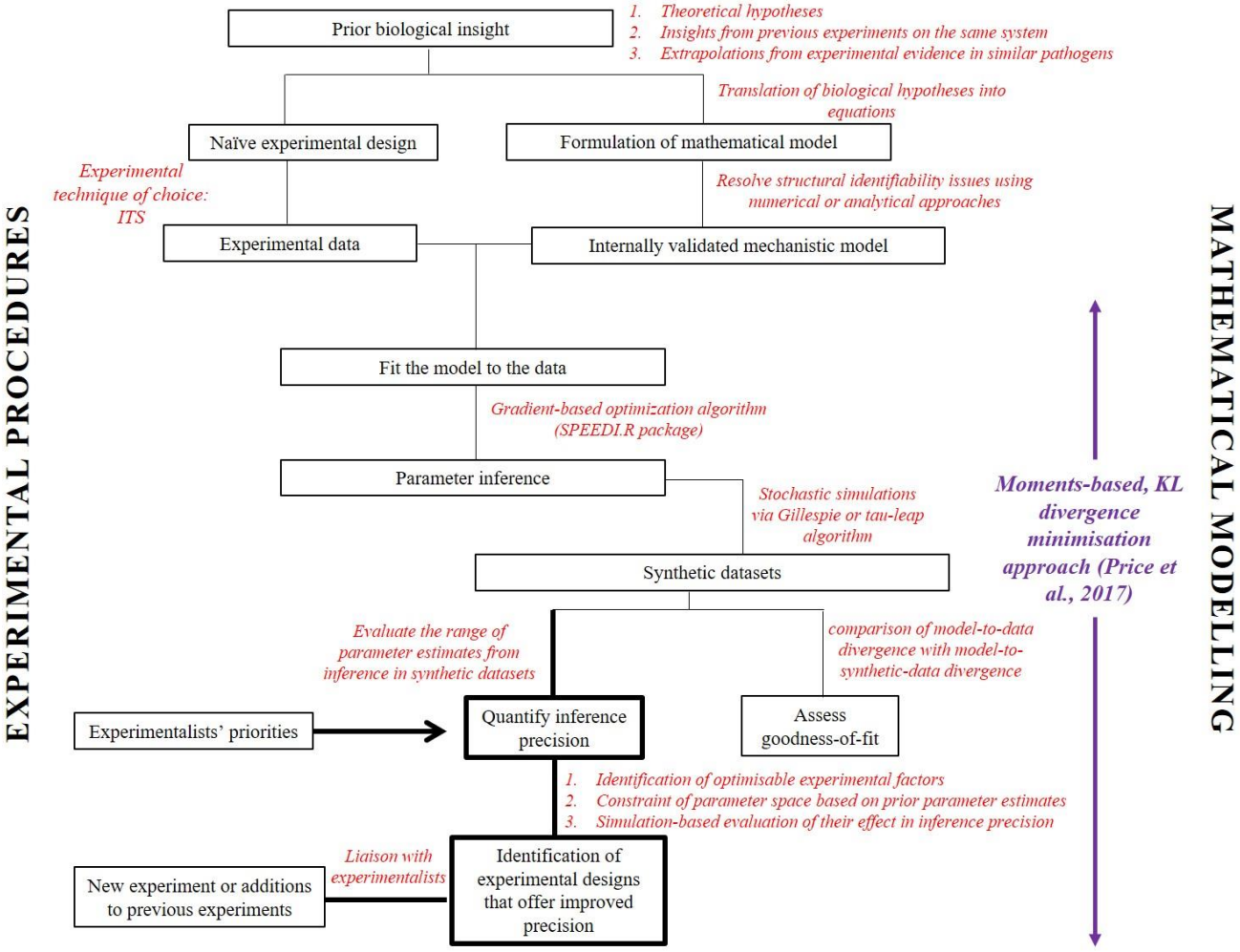


Figure 6.1: Model-informed experimental design in an integrated framework of experimental procedures and mathematical modelling.

In this paper, we extend the inference methodology to improve experimental designs and provide a case study as a blueprint for how it can be applied to data from ITS-based studies. As ITS remains one of the key experimental techniques in the field of within-host bacterial dynamics (reviewed in Vlazaki, Huber and Restif, 2019), the development of dedicated tools to facilitate the integration of mathematical modelling along the experimental process can maximise the biological insights we gain from the studies by improving the statistical quality of inferences for a substantial body of research. A unique feature of ITS-based studies is the multiplicity of infections with different tagged strains within each host organism. Under the assumptions of independent action for bacteria and homogeneity in dynamics both between ITS and host organisms, the number of biological replicates is equal to the number of ITS per host multiplied by the number of hosts. Experiments produce measurements of the number of copies of each ITS in anatomical compartments of interest at the time that each host animal is killed, and inference about the within-host dynamics "fills in" the gap between the sampled timepoints.

Taking these features of ITS infections into account, we provide the first application of mathematical modelling as a tool to improve experimental design in studies of *in vivo* bacterial population dynamics using ITS. These studies commonly address questions around the dynamical composition of bacterial populations jointly shaped by environmental and biological processes such as bacterial replication, death, migration, phenotypic switch and bacterial-host interactions. We model bacterial kinetics with a continuous time Markovian process where the state variables are the number of copies of a single WITS in each anatomical compartment. Given an experimental design, the model predicts the dynamics of the lower moments (mean, variance and covariance) of the state variables. Parameter inference is based on minimising the Kullback-Liebler divergence between the predicted and observed moments (the latter being calculated from biological replicates, i.e. all ITS from all mice at each time point). First, we formulate a utility function inspired by the moments-based, divergence minimisation approach (Price *et al*., 2017) that represents our criterion for optimality; the objective is to maximise this function across the design space. As we are interested in maximising global parameter inference precision, we focus on optimising the determinant of the parameter variance-covariance matrix. We use synthetic data to study its performance against the width of the corresponding 95% confidence intervals for a range of parameter sets. Finally, we illustrate its application by considering experimental designs for a variety of biological scenarios reflecting different phases of

bacterial dynamics including exponential growth and decay as well as cases dominated by unidirectional inter-tissue bacterial migration.

6.2. Methods

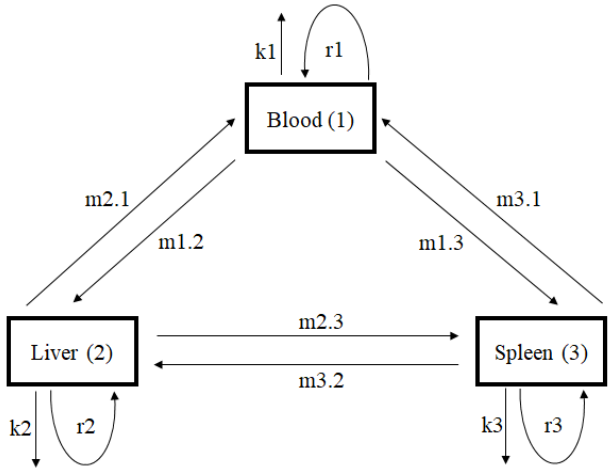
6.2.1. The ITS technique: experimental output and dataset structure

ITS represent a large class of marker-based techniques, developed over the last decade, in order to study the within-host bacterial dynamics at a population level. Libraries of isogenic tagged strains are generated by uniquely modifying the non-coding genome of bacteria with distinctive, short nucleotide sequences; the resulting ITS are genetically distinguishable, yet phenotypically identical. The ITS are mixed in equiproportional inocula and administered to the animal host. Animals are culled at predetermined time points to harvest or partly sample tissues of interest, and a total bacterial count is obtained per tissue (Abel *et al*., 2015; Grant *et al*., 2008). The samples are also processed to determine the ITS composition in each tissue, originally by qPCR (e.g. in Coward *et al*., 2014) and more recently by next-generation sequencing (e.g. in Rossi *et al*., 2017; Lim *et al.*, 2014) The final structure of the dataset consists of number of copies per ITS in each tissue of interest per time point.

By obtaining snapshots of spatiotemporal data recapitulating the changing ITS composition of bacteria in different tissues, it is possible to characterise the unobserved processes that underpin the dynamics of the infection *in vivo*. Supplementing these data with compartmental mechanistic models, the rates at which these unobserved dynamical processes evolve can be quantified with optimisable precision. We proceed to show how the inference precision can be optimised for this type of experimental data, by quantifying the impact of a range of design factors that we intuitively expect to affect inference precision.

6.2.2. Conceptual Model Structure

Mechanistic models applied to ITS data typically have a compartmental structure, with each compartment representing a tissue of interest, and model parameters corresponding to the rates at which the unobserved biological processes change with time. In this paper, we use a radial model structure with three compartments, inspired by the model used in *Salmonella* Typhimurium infections (Price *et al*., 2017; Coward *et al*., 2014; Grant *et al.*, 2008), and shown as an example in Figure 6.2. Linear compartmental model structures or a combination are also possible, and the model can be generalised to include n compartments (Price *et al*., 2017).



Parameter Name	Parameter Description
m1.2	Bacterial migration from the blood to the liver
m1.3	Bacterial migration from the blood to the spleen
m2.1	Bacterial migration from the liver to the blood
m3.1	Bacterial migration from the spleen to the blood
m2.3	Bacterial migration from the liver to the spleen
m3.2	Bacterial migration from the spleen to the liver
k1	Bacterial death in the blood
r1	Bacterial replication in the blood
k2	Bacterial death in the liver
r2	Bacterial replication in the liver
k3	Bacterial death in the spleen
r3	Bacterial replication in the spleen

Figure 6.2: Conceptual structure of the compartmental model detailing the intra- and inter-tissue dynamics in a model of systemic salmonellosis. In each organ *i*, bacteria replicate and are killed at rates r_i and k_i, respectively, and bacteria migrate to organ *j* according to rates $m_{i,j}$.

6.2.3. A likelihood-free inference framework for complex dynamical systems

As we are usually interested in tracking the evolution of infections over a period of days, and as bacterial infections are typically multi-organ, data collected from ITS-based experiments are often complex. Their complexity lies in: (1) the length of observational time for the biological system in question, (2) the number of tissues of interest, and (3) the multiple organ-organ interactions. The ITS composition across all tissues of interest is summarised in a vector of moments (means, variances, and covariances) corresponding to each observation time. Assuming first-order dynamics for each time interval between sequential points of data collection, a simple mathematical expression that derives the moments at a future time point can be obtained and solved analytically. The moments of the experimental dataset and the calculated moments for the same time point are compared via the Kullback-Leibler divergence. Minimisation of this divergence via standard optimisation routines allows us to quantify processes underpinning the change in bacterial numbers and composition during a time interval of interest. A full set of computational tools and examples on how to use this inference framework can be freely accessed at: https://github.com/orestif/SPEEDI.R.

6.2.4. Parametric bootstrap to evaluate inference precision

Parameter inference would not be complete without a measure of its precision. Quantification of precision in this framework is carried out using a parametric bootstrap approach previously described in Price *et al*. (Price *et al*., 2017). Having identified the best parameter estimate by minimising the KL-divergence between the experimental and calculated moments for a range of parameter sets, we use it to simulate 200 synthetic datasets of the same size and structure as the experimental data. The choice of 200 bootstrapped samples was motivated by our priority to maximise the computational efficiency of the suggested tool, and is in line with previous findings on the minimum adequate number of bootstrapped samples (Pattengale *et al*., 2010). The corresponding moments are calculated, and the best parameters inferred via the same divergence-minimisation method. This process results in a range of parameter estimates, which provides us with a measure of precision regarding the inference for each parameter.

6.2.5. Deriving optimality criteria in a likelihood-free inference framework

We now turn to formulating an optimality criterion in this likelihood-free framework, inspired by the gold-standard alphabetical optimality criteria in the maximum likelihood framework (Box, 1982).

In the maximum likelihood estimation (MLE) context, the Fisher Information Matrix (FIM) plays a key role in assessing parameter inference precision (Chaloner and Verdinelli, 1995). It summarises the amount of information contained in the experimentally observed dataset with regards to the parameter set $\theta = \{\theta_1, \theta_2, \theta_3, ..., \theta_n\}$. The FIM can be analytically derived from the Hessian matrix ($H(\theta)$), a matrix of 2nd order derivatives of the likelihood $L(\theta)$ with respect to the parameters:

$$FIM = -E[H(\theta)] = -E\left[\frac{\partial^2 \ln L(\theta)}{\partial\theta\, \partial\theta'}\right] \quad (1)$$

For a given MLE, the FIM contains information about the curvature of the likelihood. A higher value in the corresponding entry in the FIM matrix, indicates that the estimate for that is more precise (which can be expressed as a shorter confidence interval). Furthermore, by assessing different functions of the FIM (e.g., the determinant or trace) it is possible to

derive information about different aspects of the inference quality provided by each experimental design; these ways of assessing different aspects of the inference quality of a model given an experimental design are known as alphabetical optimality criteria (Box, 1982). For example, the determinant of the FIM provides an index of average size (e.g., width, area, volume) of the confidence interval for the parameters of interest (D-optimality criterion), while the trace of the FIM summarises the average variance of each parameter of interest (A-optimality criterion) (Box, 1982). Other composite functions of the FIM lead to the full range of the alphabetical optimality criteria.

An important property of the FIM is its inverse relationship with the variance-covariance matrix of θ, via the Cramer-Rao lower bound (Casella, 2002). In 2.3. we used a parametric bootstrap approach to estimate the variance-covariance matrix of the parameter estimates; let that be M. As the inverse of M is proportional to the FIM, we can define equivalent optimality criteria by minimising functions of M, rather than maximising functions of the FIM. Analytical evaluation of the FIM is often difficult or impossible (Spall, 2005), but instead typically approximated numerically as we propose to do with the minimum-divergence inference method. Here, we focus on minimising the determinant of M, which serves as the global measure of inference precision.

6.2.6. Identifying experimental designs with higher information content

The FIM summarises the information content of an experiment as a function of both the model parameters θ, and the design conditions under which the experiment is conducted. As such, it can be used to identify experimental designs that yield data with a higher information content. To achieve this, we first identify which aspects of the experimental process we are interested in optimising. Here, we focus on the number of biological replicates, the initial number of copies per ITS and the choice of sampling timepoints.

Next, we place sensible bounds on the parameter space to be explored. This can be done through expert elicitation, based on prior literature or understanding of the system, or by using inferences made on previous experiments on the same or a similar system.

For the scenarios we consider in Sections 6.3.1, 6.3.2.1 and 6.3.2.2, we randomly draw parameters from a uniform distribution $\theta \sim U(0,1)$. In 6.3.2.3, we explore 6 biological

patterns of within-host dynamics commonly occurring in *in-vivo* bacterial infections, which we list below. For each of the scenarios A1, A2, B1, B2, C1, C2 and keeping the total number of replicates fixed at 500, we consider 10 parameter sets, and for each parameter set 3 numbers of copies per ITS in {10,100,1000} and 8 sampling time points reflecting early, mid-, late and a combination of early and late sampling in:

t∈{(1,4),(2,5),(10,13),(11,14),(19,23),(20,24),(1,23),(2,24)}.

To provide an example of how these sampling strategies would be implemented for a time interval of 24 hours, a group of animal hosts would be euthanised at 1 hour and a separate group at 4 hours post-inoculation (1,4). In the next experimental design, a group of animal hosts would be euthanised at 2 hours and a separate group at 5 hours post-inoculation (2,5).

A. Exponential growth (net growth = +0.2, minimal inter-tissue migration)

A1. Exponential growth with high killing and high replication rates (higher variance in data)

$$k_i \sim U(2,2.5), r_i = k_i + 0.2$$

A2. Exponential growth with low killing and low replication rates (lower variance in data)

$$k_i \sim U(0.1,0.6), r_i = k_i + 0.2$$

B. Exponential decay (net growth = -0.2, minimal inter-tissue migration)

B1. Exponential decay with high killing and high replication rates (higher variance in data)

$$k_i \sim U(2,2.5), r_i = k_i - 0.2$$

B2. Exponential decay with low killing and low replication rates (lower variance in data)

$$k_i \sim U(0.3,0.8), r_i = k_i - 0.2$$

C. Source-sink unidirectional bacterial transfer with near-zero intra-organ growth

C1. Source-sink with high killing and high replication rates (higher variance in data)

$$k_i, r_i \sim U(2,2.5),\ m_{1j} \sim U(1.5,2)$$

C2. Source-sink with low killing and low replication rates (lower variance in data)

$$k_i, r_i \sim U(0.1,0.6),\ m_{1j} \sim U(1.5,2)$$

6.3. Results

6.3.1. A simulation study to correlate the determinant of the parameter variance-covariance matrix with precision in parameter inference

In this section we perform a simulation study to demonstrate that the magnitude of the determinant of the parameter variance-covariance matrix is a reliable summary statistic for the precision of inference across the entire parameter set. We start with the model structure outlined in 6.2.2 and generate 10 random parameter sets. We then simulate 10 synthetic datasets for each parameter set (10x10 = 100 in total); the synthetic datasets serve as surrogates for experimentally observed ITS-based data. For each of the 100 virtual datasets, we obtain 200 bootstrapped samples and proceed to obtain 200 sets of parameter estimates. For each set of parameter estimates, we calculate the variance-covariance matrix and obtain its determinant, which is taken to be a global measure of inference precision. Next, we obtain the 95% confidence interval for each of the 12 parameters across the 10 synthetic datasets generated for each parameter set index and report the mean of their widths across each parameter set relative to the maximum one, as a function of the corresponding determinant (Figure 6.3).

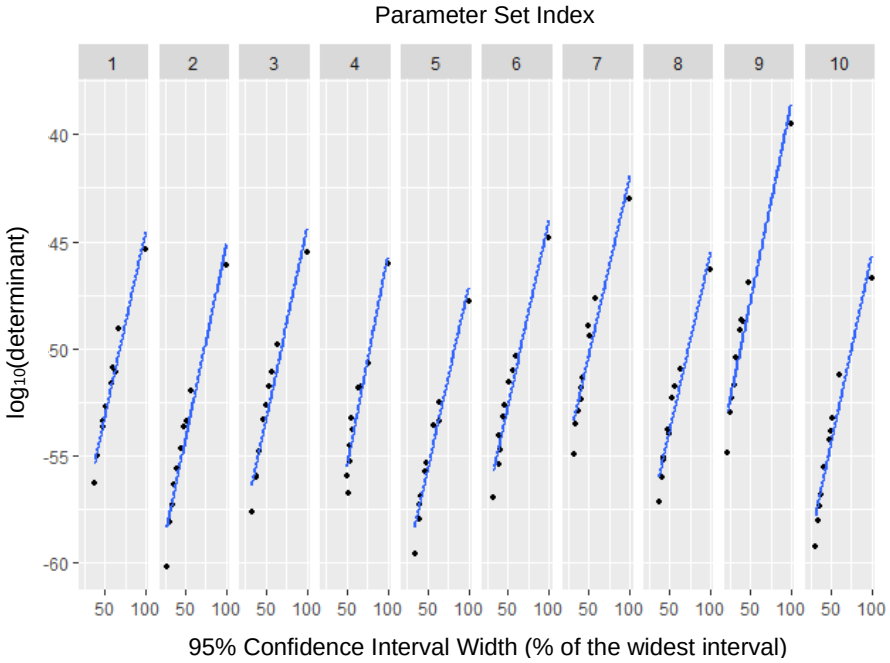


Figure 6.3: Mapping the determinant of the variance-covariance matrix to the range of 95% confidence intervals.

Each dot represents the mean width of the 95% confidence intervals across the 12 parameters in each synthetic dataset. A three-compartmental model is used (section 6.2.2) and parameters are drawn at random from a uniform distribution between 0 and 1, $\theta \sim U(0,1)$. For each of the randomly generated parameter sets (parameter set index 1-10), there are 10 dots representing 10 synthetic datasets simulated from that parameter set. The mean width of the 95% confidence intervals is plotted against the mean logarithm of the determinant across the 10 synthetic datasets generated by each of the parameter sets, showing a linear positive relationship between the two quantities.

6.3.2. Determinants of precision inference in experimental design

Having established the determinant of the parameter variance-covariance matrix as a reliable correlate to inference precision in 6.3.1, here we study the effect of 3 design factors on the quality of inference precision in ITS-based experiments, namely: the number of biological replicates, the number of copies per ITS and the choice of sampling time points. First, in 6.3.2.1, we focus on the number of biological replicates (R); under the assumption of homogeneous inter-host immunological responses (Grant *et al*., 2008; Kaiser *et al*., 2013; Coward *et al*., 2014; Rossi *et al*., 2017) R is equal to the number of animal hosts multiplied by the number of ITS per host. In 6.3.2.2, we introduce the biological limitation of a constraint on the total inoculum size (I) per host; I is equal to the number of ITS per host multiplied by the number of copies per ITS. Finally, in 6.3.2.3, we study the mixed effect of experimental designs with different numbers of copies per ITS and different sampling timepoints.

6.3.2.1. Number of biological replicates (R)

In 6.3.2.1, we use the same simulation-based approach and parameter sets as in 6.3.1. to evaluate the effect of the number of biological replicates on the magnitude of the determinant of the parameter variance-covariance matrix. At this stage, the number of biological replicates (R) is equal to the animal hosts multiplied by any number of ITS per host, without constraint on the total inoculum size per animal host. As shown in Figure 6.4, there is a directional effect between the determinant of the parameter variance-covariance matrix generated by parametric bootstrap and the total number of replicates. As the number of total replicates increases (x-axis), the quality of parametric inference increases (reflected by lower values in the determinant on the y-axis).

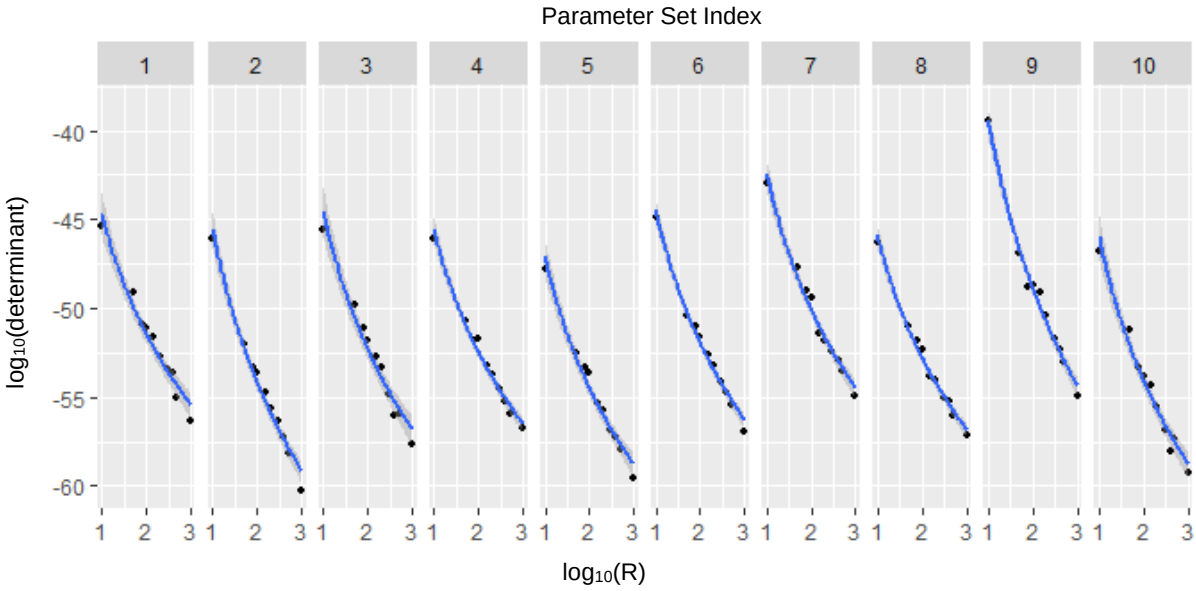


Figure 6.4: Inference precision as a function of total number of replicates.

For each parameter set, 8 experimental designs are considered with number of biological replicates $R \in \{5,10,20,50,100,200,500,10000\}$. The $\log_{10}$ of the determinant of the parameter variance-covariance matrix is plotted against $\log_{10}$ (R) to yield a consistent pattern of exponential decay across the 10 parameter sets. A three-compartmental model is used (section 6.2.2) and parameters are drawn at random from a uniform distribution between 0 and 1, $\theta \sim U(0,1)$.

6.3.2.2. Distribution of replicates in animal hosts with a constraint on total inoculum size per host (I)

Here, we move a step further from 6.3.2.1 and introduce the biological constraint of total inoculum size per host (I). In ITS-based experimental infections, there is a target total inoculum *I* achieved by multiplying the number of different ITS used by the number of copies per ITS. As detailed in Table 6.1, we consider experimental designs with the total inoculum size fixed at 10,000 CFU; this is consistent with the orders of magnitude of inocula used in previous ITS studies (Rossi *et al.*, 2017; Coward *et al.*, 2014; Kaiser *et al.*, 2013; Grant *et al.*, 2008). We also keep the number of biological replicates fixed at 1000 biological replicates to minimise the stochastic noise in experiments with small replicate sizes, and consider 8 possible combinations between the total number of host animals per sampling time point and the total number of ITS per host animal.

			Mean number of CFU per ITS				
Experimental Design Index	Total number of host animals per sampling time point	Total number of ITS per host animal	Comp 1	Compt 2	Comp 3	Total number of replicates (identical simulations)	Total inoculum size
1	200	5	2000	1	1	1000	10000
2	100	10	1000	1	1	1000	10000
3	50	20	500	1	1	1000	10000
4	20	50	200	1	1	1000	10000
5	10	100	100	1	1	1000	10000
6	5	200	50	1	1	1000	10000
7	2	500	20	1	1	1000	10000
8	1	1000	10	1	1	1000	10000

Table 6.1: A matrix tabulating the 8 experimental designs arising from different combinations of total number of host animals per sampling time point and total number of ITS per host animal, with fixed total number of replicates and fixed total inoculum size. Fixed variables are highlighted in grey, and variable design factors in yellow.

In Figure 6.5, we plot the determinant of the parameter variance-covariance matrix as a function of the total number of ITS per host animal. There is a consistent positive relationship between the two variables, indicating that lower numbers of ITS per host yield lower determinant values, which reflect higher quality inference in terms of precision. Our analysis shows that given a fixed number of biological replicates and total inoculum size per host, experimental designs with higher numbers of hosts and lower numbers of ITS per host are predicted to yield higher precision estimates.

Figure 6.5: Inference precision as a function of number of ITS per host.

For each parameter set, 8 experimental designs are considered with number of ITS per host in $\{5,10,20,50,100,200,500,10000\}$. The $\log_{10}$ of the determinant of the parameter variance-covariance matrix is plotted against the number of ITS per host to yield a consistent pattern of exponential growth across the 10 parameter sets. A three-compartmental model is used (section 2.2) and parameters are drawn at random from a uniform distribution between 0 and 1, $\theta \sim U(0,1)$.

6.3.2.3. The choice of sampling time points in designs with different numbers of copies per ITS

In 6.3.2.2. we found that given the presence of biologically realistic constraints on total inoculum size per host and total number of replicates, infections with fewer ITS at higher copy number per ITS consistently perform better in terms of precision inference. As a result, achieving high precision comes with the cost of using more animal hosts. In this section, we introduce a third design factor, the choice of sampling time points during the infection process. We seek to address the question: can we choose sampling time points that maximise the precision inference, allowing us to discount on the number of copies per ITS and, by extension, the number of animal hosts we need to inoculate?

To answer this question, we consider 6 biological scenarios that encompass common patterns in within-host dynamics of bacterial disease, described in 6.2.6. In particular, we look at scenarios of exponential decay and growth with either high or low variance, as well as cases with unidirectional bacterial migration from one organ to the other.

We start with sampling at a single observation point. In the likelihood-free inference framework, inference using a single observation time point is only possible for models whose number of parameters is smaller than or equal to the number of moments that summarise the ITS distributions. We first consider a simple, two-compartment model structure with unidirectional bacterial flow from the first to the second compartment, and bacterial death and replication in both (5 parameters in total). Figure 6.6 shows that unlike in the scenarios of exponential growth, in the scenarios of exponential decay intermediate sampling times (12- and 13-hours post-inoculation) yield higher-precision inference, whereas early or late sampling times provide less information on average. This is likely due to either minute changes in the bacterial population in the very early stages (1-2 hours), or bacterial populations having died out during the later stages (23-24 hours), at which point we can only get bounds on the parameter estimates that are consistent with extinction "up to" that time (Figure 6.6A). This pattern differs in the growth example (Figure 6.6B), where early time points will be subject to greater variation in the population sizes. Later time points will have allowed sufficient time for the populations to establish/replicate whereby the variation between subpopulations will be less impactful on the inference.

In Figure 6.7, we present the results of the same analysis performed using a combination of sampling time points for a three-compartment model. The precision across all experimental designs considered in Figure 6.7 is consistently higher than in Figure 6.6, as combining data two observation times increases the information content overall. Finally, experimental infections with higher number of copies per ITS perform consistently better in terms of inference precision (Figures 6.6 and 6.7). Assuming a constraint on the total inoculum size per host, this translates into designs involving a larger number of hosts per sampling time point.

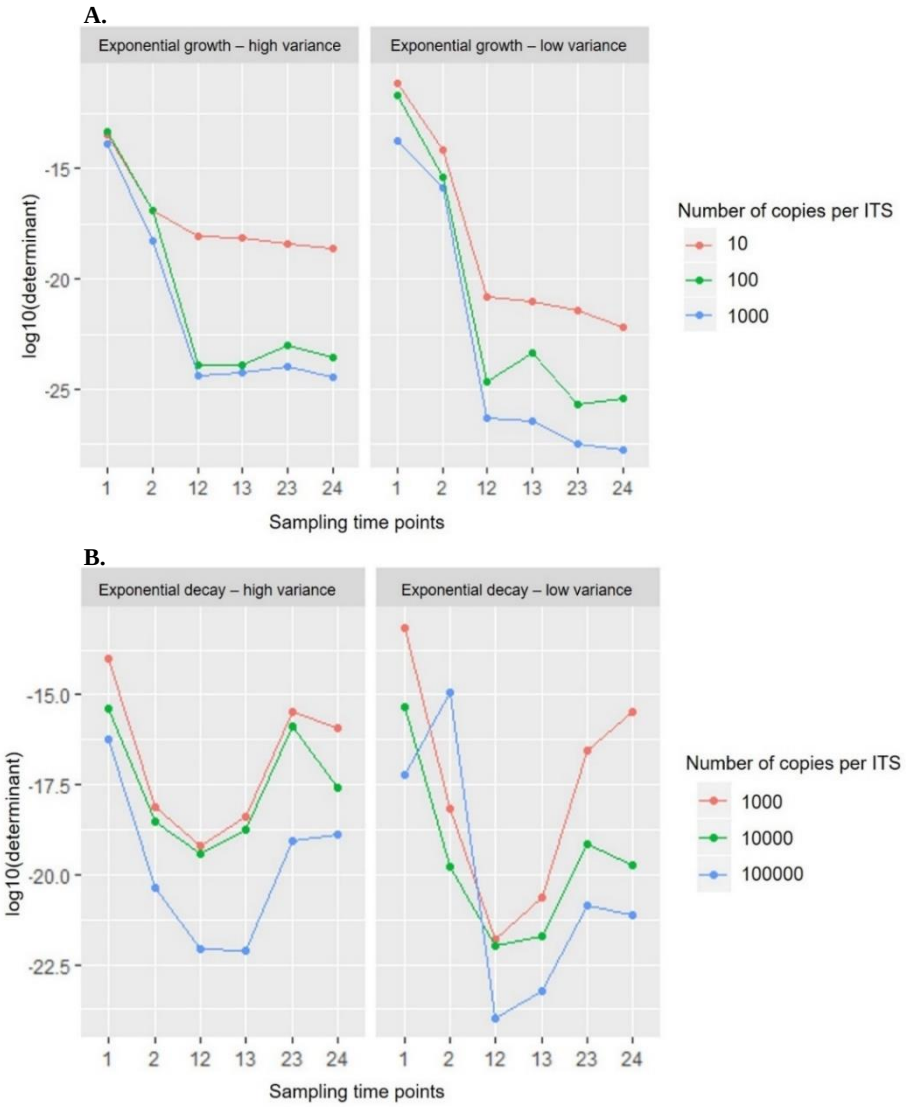


Figure 6.6: Scatter plots showing the mean $\log_{10}$(determinant) across 10 parameter sets under 4 biological scenarios with single sampling time points for a three-compartment model.

For the exponential growth scenarios, later sampling yields the highest-precision inference (Panel A). For the exponential decay scenarios, there is a U-shaped relationship between the time of sampling and inference precision, with intermediate sampling time points being the most informative. A two-compartmental model with intra-organ bacterial growth and death, and bidirectional bacterial migration is used. Parameters are drawn at random from a uniform distribution between 0 and 1, $\theta \sim U(0,1)$.

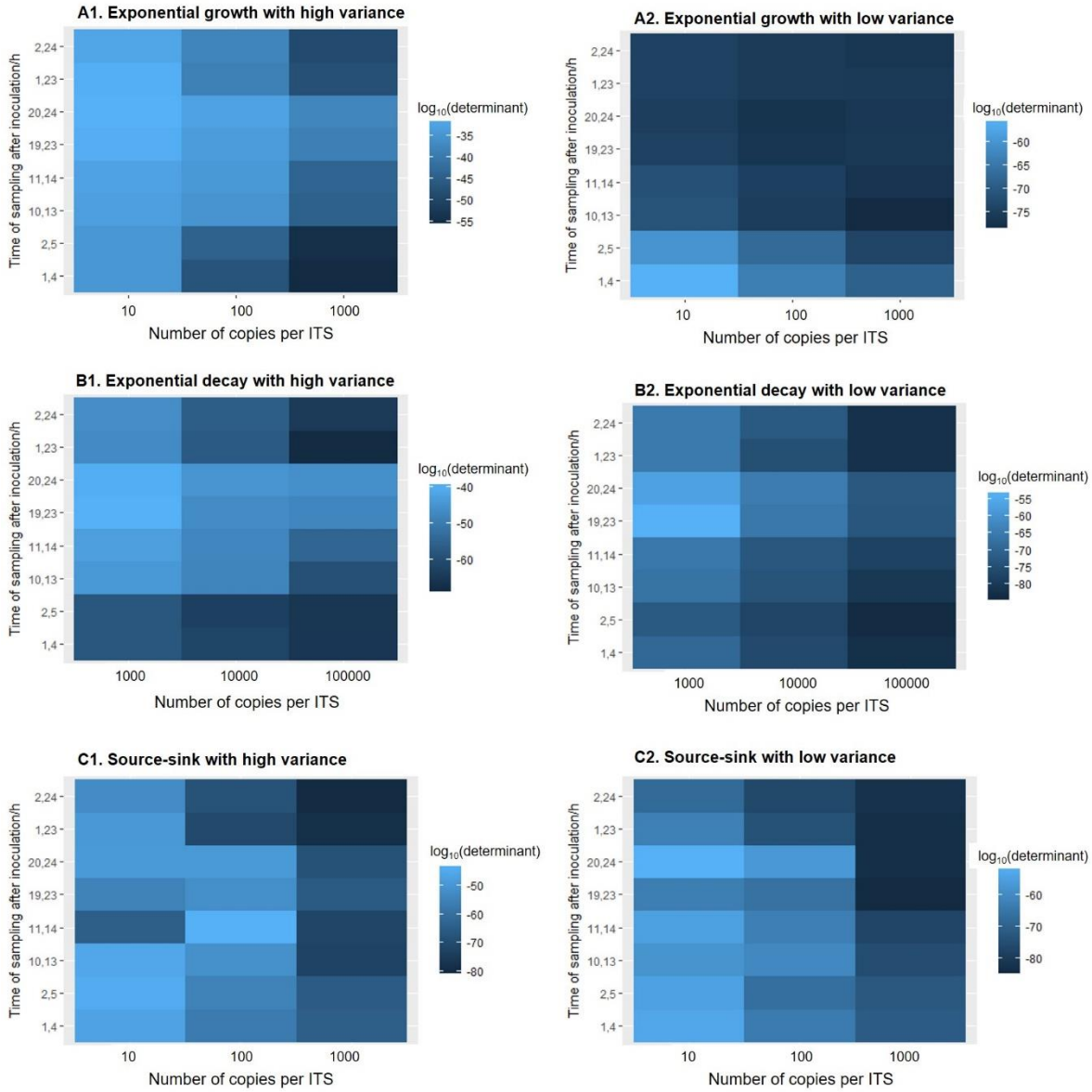


Figure 6.7: Heatmap plots showing the mean $\log_{10}$(determinant) across 10 parameter sets under 6 biological scenarios for a range of experimental designs with variable sampling time points and starting number of copies per ITS for a four-compartment model.

The first column of panels (A1, B1, C1) illustrates biological scenarios with high killing and replication rates, while the second column (A2, B2, C2) corresponds to biological scenarios with low killing and replication rates. Darker blue shades represent experimental designs yielding higher inference precision, while lighter blue shades represent designs with lower inference precision. For all experimental designs across all scenarios (A1-C2), higher numbers of copies per ITS improve the precision inference. For low-variance scenarios (low replication, low killing), the inference precision is globally better and the number of high precision experimental designs is larger than for high-variance scenarios. A three-compartmental model is used (section 6.2.2) and parameters are drawn at random from distributions specified in section 6.2.5.

6.3.3. Strategies to reduce the number of animal hosts

In 6.3.2.3, we showed that infections with higher numbers of copies of ITS per host tend to correlate with higher inference precision consistently across the sampling time points we considered. Under a fixed inoculum size per host, this translates into the cost of requiring more animal hosts, which incurs primarily ethical, but also financial costs. The analysis in 6.3.2.3 assumed a fixed number of biological replicates of 500.

Here, we consider the interaction between the total number of biological replicates (number of animal hosts multiplied by number of ITS per host) and the number of ITS per host, the latter being inversely correlated with the initial number of copies per ITS. Taking the case of a system with exponential growth dynamics and high variance as an example, we consider the experimental designs tabulated in Table 6.2. Each of A to C include experimental designs with a decreasing number of animal hosts per timepoint. From our previous analysis, the first design in each category A-C is expected to yield the highest precision, as it requires a high number of animal hosts and a high number of initial copies per ITS. In alternative designs 2 and 3 there is a ten-fold decrease in the number of animal hosts, with design #2 having a higher number of biological replicates combined with a low number of copies per ITS and design #3 having a low number of biological replicates combined with a high number of copies per ITS.

			Mean number of CFU per ITS				
Experimental Design Index	Total number of host animals per sampling time point	Total number of ITS per host animal	Compartment 1	Compartment 2	Compartment 3	Total number of biological replicates (identical simulations)	Total inoculum size
A1	200	5	2000	1	1	1000	10000
A2	20	50	200	1	1	1000	10000
A3	20	5	2000	1	1	100	10000
B1	100	5	2000	1	1	500	10000
B2	10	50	200	1	1	500	10000
B3	10	5	2000	1	1	50	10000
C1	50	10	1000	1	1	500	10000
C2	5	100	100	1	1	500	10000
C3	5	10	1000	1	1	50	10000

Table 6.2: Experimental designs to evaluate the relative effects of two design factors (total number of replicates and initial number of copies per ITS) on inference precision.

In Figure 6.8, we comparatively plot the mean $\log_{10}$(determinant) for the 3 classes of experimental designs outlined in Table 6.2, across a range of sampling time points. While experimental design index 2 and 3 use the same number of animal hosts, the latter performs significantly better and very close to experimental design index 1, which involves 10 times more animal hosts. Overall, our analysis shows, that given a constraint on the number of animal hosts, higher inference precision is achieved by using experimental infections with large initial numbers of ITS even at lower number of biological replicates; the effect of the initial number of ITS is stronger than that of number of biological replicates in determining inference precision.

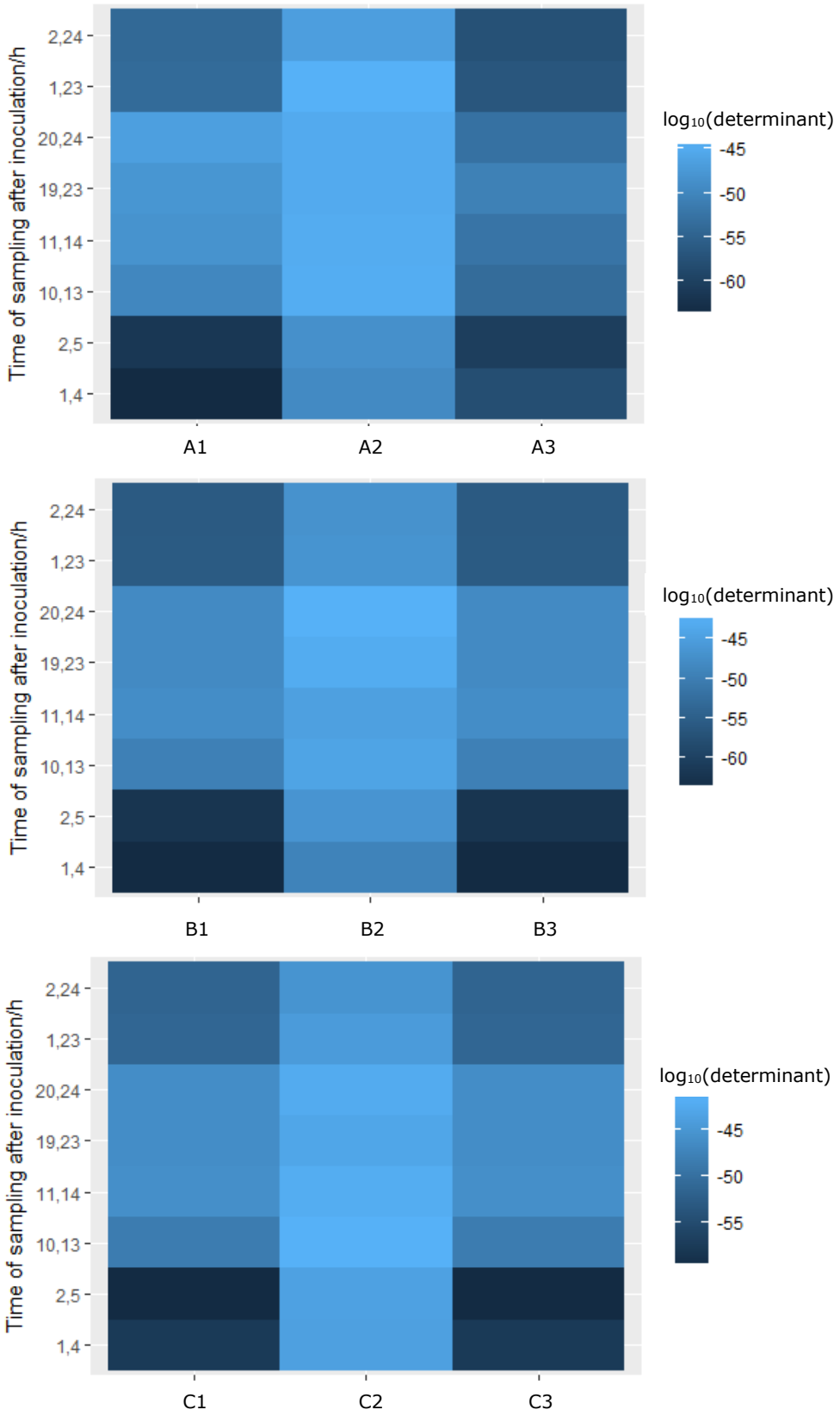


Figure 6.8: Heatmap plots showing the mean $\log_{10}$(determinant) across 10 parameter sets for experimental designs 1-3 in classes A-C with variable number of biological replicates and initial number of copies per ITS.

6.4. Discussion

We have presented a novel, computationally efficient approach to information-based experimental design for ITS-based data, with the aim of maximising inference precision for the within-host dynamics of bacterial infections. Inspired by the canonical D-optimality criterion, we chose the determinant of the parameter variance-covariance matrix, obtained via a parametric bootstrap, as a criterion for global parameter inference precision in a likelihood-free inference framework previously developed by Price *et al.* (Price *et al.*, 2017*)*. We used a simulation study to illustrate that the magnitude of the determinant effectively captures the mean width of confidence intervals across parameter sets. As the determinant is a global measure of variance, we correlated it with the mean width of the confidence interval across the entire parameter set (12 parameters in our model) and demonstrated the strong correlation between the value of the determinant and the average width of the confidence interval.

We proceeded to show that the replicate size, initial number of copies per ITS and sampling timepoint selection all have an impact on inference precision in range of general biological cases representative of bacterial dynamics typically observed during experimental infections. Biological scenarios with low replication and killing rates are characterised by low variance in the distribution of copies per ITS (Price *et al.*, 2017*)*; for the entirety of experimental designs considered for these low-variance scenarios, the inference precision is globally higher than for higher-variance scenarios. Furthermore, there is greater redundancy in experimental designs yielding similar levels of high-precision inference for lower-variance scenarios, entailing that prior intuitive knowledge about the dynamics of a system can inform experimental microbiologists about how much focus they should place on identifying unique experimental designs with the highest information content. Finally, we found that for a fixed total number of replicates and assuming inter-host immunological homogeneity, infections with lower numbers of ITS per host at higher initial copies per ITS tend to give higher-level inference precision for both high- and low-variance scenarios.

The benefit of our approach is that it can be used not only to identify a single optimal experimental design, but also to rank experimental designs, thus providing flexibility to experimentalists to choose those that reflect their own priorities. For example, assuming that the reduction in the number of animal hosts is a priority, one can dismiss the globally optimal

experimental design that requires using a 10-fold higher number of animal hosts and can choose the design with the second highest precision, which can be achieved by strategically choosing the sampling time points while keeping the number of animal hosts low. Finally, we looked at the interaction between two design factors (total number of replicates and initial copies per ITS) under a constraint on number of animal hosts, which is one of the key priorities in terms of animal welfare, logistical and financial considerations alike. We show that in this scenario, favouring an experimental design with lower number of biological replicates and fewer ITS per host yields consistently higher inference precision than the alternative allocation of resources with high number of biological replicates and more ITS per host. This consideration becomes particularly relevant as individual ITS detection in a multi-ITS sample has become more advanced, moving away from qPCR with a detection threshold of a <10 ITS per host tissue (Kaiser *et al*., 2013; Grant *et al*., 2008) to next generation sequencing able to detect tens or hundreds of ITS per host tissue sample (Zhang *et al*., 2017; Abel *et al*., 2015).

Discussing the assumptions and limitations of the likelihood-free inference framework used as the basis for the experimental design tool presented here, the moments-based, divergence-minimisation inference framework makes the simplifying assumption that these distributions are multivariate normal across all tissues of interest and sampling time points; this assumption facilitates the derivation of the expressions for the moments. Given that the shape of the true distribution is different for every tissue and sampling time point, and that one would need to approximate each of those individually, the simplifying choice of a multivariate normal distribution seems sensible. Furthermore, the first- and second- order moments of these distributions were used as sufficient summary statistics. Although it could be argued that higher order moments would capture additional features of the distributions, the added complexity is not justified by the resolution of the available data and the previous simplifying assumption about the structure of the distributions. As shown by Price *et al*., 2017, there is excellent agreement between both the inferences and the width of the confidence intervals obtained when using the maximum-likelihood framework and the likelihood-free framework used here; as a result, the use of summary statistics and the assumption of multivariate normal distributions do not affect the quality of inference.

In terms of the bootstrapped samples, we limited their number to 200 per experimental design considered in this study, striking a balance between reducing inter-simulation

variation (Monte Carlo error) and limiting the applicability of the tool due to long running times. Earlier work by Price *et al*., 2017 shows that even at 100 experiments (or bootstrapped samples), the Gillespie and moments-based approaches both lead to more and more precise estimates as the number of biological replicates increases, while at 500-1000 biological samples, the variance in the parameter estimates converges to 0. Hence, by large, the source of the Monte Carlo error is not the number of bootstrapped samples, but the number of biological replicates, which is one of the design factors considered in this study. Furthermore, our choice of 200 bootstrapped samples was further motivated by Pattengale *et al*., 2010, who developed a computationally efficient bootstrapping algorithm in the field of phylogenetics and used it to formulate testing criteria for determining the minimum number of bootstrapped samples needed to infer confidence values; in most cases, a size of 100-500 bootstrapped samples was adequate. Finally, we considered at least 10 parameter sets for each experimental design, obtaining consistent results in terms of how the experimental designs ranked relative to each other across the 10 individual datasets; this consistent pattern further confirms that our choice of bootstrapped sample size is not a major source of Monte Carlo error.

As exploratory research is gaining momentum and the competition for funding is increasing, optimal resource allocation is emerging as a first step in the process of experimental setup conceptualisation (O'Malley, 2007), while also affecting the welfare of animal hosts which is further reflected in the principle of the 3Rs – replacement, reduction and refinement (Russell and Burch, 1959; Sneddon, Halsey and Bury, 2017). *In vivo* studies in population dynamics microbiology have typically not used experimental design tools until now. For example, despite the sustained wide application of the ITS technique, questions around optimising experimental design and resource allocation have remained formally unaddressed, with the exception of past studies having used small-scale simulation-based approaches to determine a suitable inoculum size (Coward *et al*., 2014) or distribution of ITS per host (Coward *et al*., 2014; Kaiser *et al*., 2013). Limited examples of efforts to develop experimental design tools for population-level data in dynamical systems can be found in other microbiological fields like systems biology (Braniff, Richards and Ingalls, 2019; Sinkoe and Hahn, 2017; White *et al*., 2016; Mdluli, Buzzard and Rundell, 2015; Liepe *et al*., 2013, Steiert *et al*., 2012), ecology (Moffat *et al*., 2020; Zhang *et al*., 2018), and food science (Garre *et al*., 2019; Price *et al*., 2018).

The apparent under-representation of optimal design algorithms in population dynamics microbiology is largely attributable to the lack of computationally efficient inference methods that cater for mathematical models with intractable likelihoods. Such models are often necessary to capture even simplified representations of biological systems which become increasingly more complex, as infections progress (Price *et al*., 2017). Here, we have provided a blueprint for the use of experimental design tools in complex systems in population dynamics microbiology. We addressed replicate size, sampling time point selection, and the initial number of copies per ITS in the inoculum as design factors commonly considered by experimentalists when trying to optimise multifactorial resource allocation to achieve a desired level of inference precision. Our framework for optimising experimental design can be further extended to serve a wider range of experimentalists' goals beyond maximising inference precision. Other common goals include improving experimental designs to enable model discrimination, which can aid in assessing competing biological hypotheses (Vanlier *et al*., 2014). In this paper, we focused on improving the overall inference precision and focused on the determinant of the parameter variance-covariance matrix as the utility function of choice. However, since our likelihood-free experimental design tool is based on an approximation of the FIM, it could accommodate other utility functions such as the trace or the eigenvalue of the parameter variance-covariance matrix, which correspond to the A- (Chernoff, 1953) and E- (Ehrenfeld, 1955) optimality criteria, respectively.

7.

Discussion and Future Directions

7.1. Synthetic Overview

Overall, the present thesis constitutes an interdisciplinary effort to explore both the applied and methodological aspects of mechanistic mathematical models when applied to data from ITS studies at the level of the bacterial population in host tissues. On the one hand, mathematical models were applied to generate biological insights, and, on the other, methodological tools were developed to further improve the inferential process. First, I developed and used mechanistic mathematical models in conjunction with experimental data to address questions regarding the within-host dynamics of *S.* Typhimurium infection in the naïve (Chapter 3), immunised (Chapter 4) and antibiotic-treated (Chapter 5) murine host. Second, on a methodological level, I used mathematical models for inference (Chapters 3-5), model selection (Chapters 3, 5) and experimental design (Chapter 6). To the best of my knowledge, this is the first time that data-based mechanistic models have been used so extensively in the field of bacterial dynamics and applied to complex biological systems over a longer infection time period (> 24 hours).

To summarise the key findings presented in this thesis, the use of mathematical models has enabled the characterisation of the unobserved within-host dynamics in the naïve, immunised and antibiotic-treated host. In Chapter 3, I showed that inter-organ bacterial mixing already occurs in the first 6 hours post-inoculation, much earlier than previously thought. The early dynamics in the liver and spleen differ; bacteria preferentially migrate to the liver where they are subject to high bactericidal activity, while the spleen microenvironment is permissive to bacterial growth. A balance between bacterial killing and replication is reached by 24 hours post-infection in both the liver and spleen. In Chapter 4, I compared the effect of a live-attenuated and whole-cell killed vaccine and showed that the superior performance of the former in preventing bacteraemia is largely attributable to its early inactivation of the secondary challenge. I then discussed the differential effects that the killed and live vaccine exerted on the within-host dynamics of *S.* Typhimurium in the

context of the significant difference in IgG2 levels induced by the two formulations. In Chapter 5, comparing the effect of ciprofloxacin and ampicillin on the within-host dynamics of *S.* Typhimurium both during treatment and after antibiotic withdrawal revealed that bacteria surviving treatment in the spleen and mesenteric lymph nodes of ampicillin-treated mice were non-replicating, while there was no evidence for the presence of the non-replicating phenotype in ciprofloxacin-treated mice. Following bacterial relapse after antibiotic withdrawal, mathematical inference suggests that previously non-replicating bacterial subpopulations grow faster than the progeny of previously replicating bacteria. Finally, in Chapter 6, I developed a computationally efficient tool for optimising experimental design in ITS-based studies, and showed that decisions on best resource allocation under constraints can be greatly aided prior to performing the experiment, leading to optimised use of animal hosts, choice of sampling points and distribution of bacterial copies per ITS.

7.2. Implications of my research: new insights, new questions

The *Discussion* section at the end of each chapter provides a focused contextualisation of the implications of the findings presented at the chapter in question. Rather than summarising these discussion points, in this section, I have chosen to discuss the broader implications of my research on a methodological level and in terms of the generated biological insights.

On a methodological level, the work presented here provides a blueprint for the extensive use of an integrated modelling-data approach, not only as a tool for *a posteriori* data analysis (Chapters 3-5), but also for *a priori* experimental design (Chapter 6). Showcasing the merits of this dual role of mathematical modelling has the potential to suggest a new paradigm in which microbiological research is carried out under conditions that maximise the information content of the generated output, in line with the priorities set by the experimentalists. Of course, this new field of vision comes with its own unchartered philosophical questions and challenges: is there a danger that an *a priori* optimisation of experimental design introduces bias in the generated output? what are the implications of sequential experimental design, and can experimental design be used iteratively to enrich existing datasets in sequential experimental designs?

At the interface between mathematical modelling and biology, inferences made on the basis of mechanistic models characterise fine-resolution mechanisms that have not yet been experimentally observed. These inferences can lead to the rejection of previous conjectures, the generation of new hypotheses and the suggestion of appropriate experimental procedures to obtain the missing insight. For example, the analysis in Chapter 2 shows that the experimentally observed patterns of bacterial distribution per ITS can only be explained by a model that includes early organ-to-blood migration; this challenges the idea that *Salmonella* populations in the systemic organs are segregated in the first phase of the infection. Greater inter-tissue bacterial mobility raises more questions as to whether these bacteria migrate as single bacteria, host cell-bound or host cell-internalised. To address this question, real-time tracking *in vivo* studies of single *Salmonellae* migrating between tissues would be required. Another example of how insights generated through an integrated data-driven modelling approach can guide further experimental work is seen in Chapter 5, which adds evidence to the hypothesis that the emergence of bacterial persisters depends on the dose, frequency and route of antibiotic administration, as well as the class of the chosen antibiotic. A growing body of knowledge around the wide range of factors affecting antibiotic persistence could lead to experimental setups that can study chosen putative determinants of interest in isolation, while carefully controlling for the rest; this could address the incongruence in conclusions reached by different studies.

Finally, the work presented in Chapters 4 and 5, which identifies bactericidal and bacteriostatic effects of therapeutic and prophylactic interventions, can inspire a new way of defining bactericidal and bacteriostatic agents. Currently, the definition of bactericidal and bacteriostatic agents is based on a combination of *in vitro* microbiological activity assays that determine the minimum inhibitory concentration (MIC), time-kill curve and serum bactericidal titre (SBT), whose results differ depending on the conditions under which they were carried out, and whose clinical interpretation is problematic (Pankey and Sabath, 2004; Vosti, 1989). Quantification of the bacterial *in vivo* replication and killing rates in the antibiotic-treated host and comparison against a naïve control group may give an alternative and more clinically relevant indication of the cidal and static effects of a given antibiotic.

In summary, the various thematic areas explored with the aid data-driven mechanistic mathematic models in the present thesis showcase that this approach can provide evidence either for or against a biological hypothesis, guide further experimental work, and provide

additional tools to characterise and benchmark the mode of action of therapeutic and prophylactic agents. Finally, developing a tool for experimental design in microbiological studies will allow for more minimal and more effective resource allocation to generate maximally informative experimental outputs.

7.3. Limitations of my research

As detailed in 7.1 and 7.2, mathematical models, especially when applied to experimental data, can generate insights which would, otherwise, go unrecognised. It is, however, imperative to summarise and highlight the limitations mathematical models are subject to, because their explicit recognition can inspire solutions, and determine their merit on realistic grounds. I will discuss the limitations of my research in terms of the modelling approach, model selection, and interpretation of the inference results.

With regards to whether the data used were fit for purpose, it is important to highlight the assumptions behind the modelling approach and how well they reflect the data structure. As noted in the *Methods* section, ITS-based data provide a static snapshot of the distribution of bacterial copies per ITS in each tissue of interest. The moments-based, divergence-minimisation inference framework used here makes the simplifying assumption that these distributions are multivariate normal across all tissues of interest and sampling time points; this assumption facilitates the derivation of the expressions for the moments. Given that the shape of the true distribution is different for every tissue and sampling time point, and that one would need to approximate each of those individually, the simplifying choice of a multivariate normal distribution seems sensible. Furthermore, in this thesis and in line with work by Price *et al.* (Price *et al.*, 2017), the first- and second- order moments of these distributions were used as sufficient summary statistics. Although it could be argued that higher order moments would capture additional features of the distributions, the added complexity is not justified by the resolution of the available data and the previous simplifying assumption about the structure of the distributions. An additional assumption made here is that bacteria behave as independent entities *in vivo*, and thus the probability of a demographic event taking place is modelled as uniform across the entire bacterial population in that tissue (Cornforth *et al.*, 2015). However, it is known that bacteria coordinate their *in vivo* behaviour population-wide via a range of quorum sensing systems, which control bacterial migration and replication in response to environmental cues such as food abundance or depletion, crowded conditions and other.

Regarding model selection (Chapters 3, 4 and 5), there are no gold-standard criteria like those in the likelihood framework (e.g. BIC and AIC). For the purposes of this thesis, I developed a conservative algorithm based on the law of parsimony, to enable selection of the model that not only fits the average global patterns in the data (lowest overall KL divergence), but also adequately approximates all individual summary statistics (moments) by evaluating the proximity of the experimentally observed value to the mean value of each moment, as determined by stochastic simulations of that model. Model selection between nested models (Chapters 3 and 5) is challenging and the marginal "value" a modeller attaches to an additional gained unit of goodness-of-fit is largely up to debate (Kuha, 2004). In this work, model selection was guided by the rule of parsimony; the simplest model that adequately aligns with the pre-chosen summary statistics was chosen as the best model.

In terms of interpreting the inferences made using this framework and the ITS-based data, the first limitation to highlight is that the reported rates represent average rates across the entire bacterial population in a tissue of interest. Heterogeneity in bacterial phenotypes is a well-documented phenomenon (Helaine and Holden, 2013) which is, nonetheless, not reflected in this analysis. Another source of heterogeneity that goes largely undiscussed in microbiological studies is the inter-host heterogeneity in immunological responses. ITS-based data reflect bacterial numbers in a pool of murine hosts, which although breed-, age- and sex-matched are, nonetheless, different organisms with potential differences in their immunological responses. Finally, the inter-sampling time intervals in the datasets used in the present thesis range from 0.5 to 24 hours. In my analysis, I assumed the simplest case of linear rates to capture the unobservable dynamics of the experimental system during these time intervals. Although it is a convenient choice because it results in much simpler mathematical expressions, it is also justified by the fact that only average rates are estimated, and that for a small time interval, more complex rate expressions may introduce the risk of overfitting.

In summary, this section highlighted the assumptions and limitations behind the modelling work presented in this thesis, in terms of the modelling approach, model selection and inference interpretation. As noted in the next section, cataloguing these limitations can be an invaluable guide for future experimental and modelling work.

7.4. Future Directions

As discussed earlier, data-driven mathematical modelling in microbiology is an interdisciplinary nascent field, whose evolution depends both on the advent of higher-resolution bacterial tracking techniques *in vivo*, and on the development of appropriate, computationally efficient modelling tools. As developments on both fronts are underway, I would like to conclude this thesis by providing suggestions for further work both in terms of methodological tools and in terms of new experiments.

With regards to methodological tools and improvements to the current inference framework, there is scope to address the challenge of inter-host immunological heterogeneity by incorporating a hierarchical modelling approach within the existing inference framework. In this approach, the within-host dynamics are allowed to differ between each host animal, are simulated under different sets of parameters, are combined, and summarised by a single set of summary statistics. The second line of work would involve developing a tool to assess model mimicry, which can be simply explained as the degree of overlap between the predictions made by two different models. To give a brief outline, this would involve two datasets (1 and 2) and two rival models (M1 and M2). M1 and M2 would be fitted to both datasets 1 and 2, resulting in four best parameter estimates. Using a parametric bootstrap approach, each of these four parameter sets would be used to simulate a large number of datasets, and the models would be fitted to all bootstrapped datasets using the KL-divergence minimisation approach, resulting in four distributions of KL-divergence values. It is, then, possible to quantify the degree of overlap between the divergence distribution of the correct model M1 applied to dataset 1, and the wrong model M2 applied to dataset 1, and similar for dataset 2. Taking the ratio of these areas of overlap, it is possible to quantify the extent of model mimicry for M1 and M2.

With regards to suggestions for further experiments, this work by large inspires more experimental work at the level of the single bacterium. As noted earlier (section 7.3), heterogeneity in bacterial behaviour has been largely unaccounted for when using data at the level of the entire bacterial population. Individual host cell-bacterial interactions are an important area to investigate when trying to understand the *in vivo* effect of antibiotics or vaccines. Furthermore, with a well-established legacy of ITS experiments relating biological mechanisms to immunological components, it would be good to see future experiments integrating the two in a single experimental setup, by e.g. sampling the blood serum for the

presence of immunological markers that may be of interest. Finally, follow-up experiments with genetically-engineered hosts can both externally validate and enrich the understanding provided by mechanistic mathematical models as shown by Grant *et al*. (Grant *et al*., 2008) and Kaiser *et al*. (Kaiser *et al*., 2013).

7.5. Conclusion

In conclusion, the present thesis highlights the merits of mechanistic mathematical models when combined with experimental data before, during and after experimentation in the lab. As a new interdisciplinary field, modelling in microbiology needs to be fostered and gain visibility by showcasing successful examples of how it can maximise the insight produced by wet-lab experiments. Besides gaining visibility, it is also critical to encourage the dialogue between modellers and microbiologists, to fund interdisciplinary projects and to promote cross-disciplinary expertise in both life sciences and applied mathematics.

8.

Bibliography

1. Abel S, Abel zur Wiesch P, Chang HH, Davis BM, Lipsitch M, Waldor MK. 2015. Sequence tag-based analysis of microbial population dynamics. Nat Methods. 12(3):223-226.

2. Abel, S., Abel zur Wiesch, P., Davis, B. M., & Waldor, M. K. 2015. Analysis of Bottlenecks in Experimental Models of Infection. PLOS Pathogens, 11(6), e1004823.

3. Angerman, C., & Eisenstein, T. 1980. Correlation of the duration and magnitude of protection against Salmonella infection afforded by various vaccines with antibody titers. Infection And Immunity, 27(2), 435-443.

4. Antia R, Levin BR, May RM. 1994. Within-host population dynamics and the evolution and maintenance of microparasite virulence. Am Nat. 144:457–72.

5. Ao, T., Feasey, N., Gordon, M., Keddy, K., Angulo, F. and Crump, J. 2015. Global Burden of Invasive Nontyphoidal Salmonella Disease, 20101. Emerging Infectious Diseases, 21(6), pp.941-949.

6. Avery, S. 2006. Microbial cell individuality and the underlying sources of heterogeneity. Nature Reviews Microbiology, 4(8), pp.577-587.

7. Avraham, R., Haseley, N., Brown, D., Penaranda, C., Jijon, H., Trombetta, J., Satija, R., Shalek, A., Xavier, R., Regev, A. and Hung, D. 2015. Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses. Cell, 162(6), pp.1309-1321.

8. Bacterial vaccines. 2020. https://www.drugs.com/drug-class/bacterial-vaccines.html [Accessed on 10/07/2020].

9. Bakaletz, L. O. 2004. Developing animal models for polymicrobial diseases. Nature Reviews Microbiology, 2(7), 552–568.

10. Bakardjiev AI, Theriot JA, Portnoy DA. 2006. Listeria monocytogenes traffics from maternal organs to the placenta and back. PLoS Pathog. 2:e66.

11. Baker RE, Peña J-M, Jayamohan J et al. 2018. Mechanistic models versus machine learning, a fight worth fighting for the biological community? Biol Lett. 14:20170660.

12. Bakkeren, E., Huisman, J. S., Fattinger, S. A., Hausmann, A., Furter, M., Egli, A., Slack, E., Sellin, M. E., Bonhoeffer, S., Regoes, R. R., Diard, M., & Hardt, W.-D. 2019. Salmonella persisters promote the spread of antibiotic resistance plasmids in the gut. Nature, 573(7773), 276–280.

13. Balaban, N. Q. 2004. Bacterial Persistence as a Phenotypic Switch. Science, 305(5690), 1622–1625.

14. Banerjee S, Keval R, Gakkhar S..2013. Modeling the dynamics of Hepatitis C virus with combined antiviral drug therapy: Interferon and Ribavirin. Mathematical Biosciences. 245:235–48.

15. Barnes PD, Bergman MA, Mecsas J et al. 2006.Yersinia pseudotuberculosis disseminates directly from a replicating bacterial pool in the intestine. J Exp Med.203:1591–601.

16. Beaumont, M. A. 2010. Approximate Bayesian Computation in Evolution and Ecology. Annual Review of Ecology, Evolution, and Systematics, 41(1), 379–406.

17. Bernatová, S., Samek, O., Pilát, Z., Serý, M., Ježek, J., Jákl, P., Siler, M., Krzyžánek, V., Zemánek, P., Holá, V., Dvořáčková, M., & Růžička, F. 2013. Following the mechanisms of bacteriostatic versus bactericidal action using Raman spectroscopy. Molecules (Basel, Switzerland), 18(11), 13188–13199.

18. Beven K, 2006. A manifesto for the equifinality thesis. J Hydrol 320, 18–36.

19. Bigger J. W. 1944. Treatment of staphylococcal infections with penicillin by intermittent sterilisation. Lancet 244 497–500.

20. Blanden, R. V., Mackaness, G. B., & Collins, F. M. 1966. MECHANISMS OF ACQUIRED RESISTANCE IN MOUSE TYPHOID. Journal of Experimental Medicine, 124(4), 585–600.

21. Bottery, M. J., Passaris, I., Dytham, C., Wood, A. J., & van der Woude, M. W. 2019. Spatial Organization of Expanding Bacterial Colonies Is Affected by Contact-Dependent Growth Inhibition. Current Biology, 29(21), 3622–3634.e5.

22. Box, G. E. P. 1982. Choice of Response Surface Design and Alphabetic Optimality. Utilitas Mathematica, 21B: 11-55.

23. Braniff, N., Richards, A., & Ingalls, B. (2019). Optimal Experimental Design for a Bistable Gene Regulatory Network. IFAC-PapersOnLine, 52(26), 255–261.

24. Brauner, A., Fridman, O., Gefen, O. and Balaban, N. 2016. Distinguishing between resistance, tolerance and persistence to antibiotic treatment. Nature Reviews Microbiology, 14(5), pp.320-330.

25. Breiman RF, Cosmas L, Njuguna H et al. 2012. Population-based incidence of typhoid fever in an urban informal settlement and a rural area in kenya: implications for typhoid vaccine use in Africa. PLoS One. 7:e29119.

26. British Medical Journal (BMJ) Best Practice. 2018. [online] Available at: https://bestpractice.bmj.com/topics/en-us/221 [Accessed on 26/07/18].

27. Broach JR, Thorner J. 1996. High-throughput screening for drug discovery. Nature. 384(6604 Suppl):14-6.

28. Brown SP, Cornell SJ, Sheppard M et al. 2006.Intracellular demography and the dynamics of Salmonella enterica infections. PLoS Biol. 4:e349.

29. Bumann, D., & Cunrath, O. 2017. Heterogeneity of Salmonella-host interactions in infected host tissues. Current Opinion in Microbiology, 39, 57–63.

30. Burgman MA, Ferson S, Akcakaya HR. Risk assessment in conservation biology. London: Chapman and Hall, 1993.

31. Cantone, M., Santos, G., Wentker, P., Lai, X. and Vera, J. 2017. Multiplicity of Mathematical Modeling Strategies to Search for Molecular and Cellular Insights into Bacteria Lung Infection. Frontiers in Physiology, 8.

32. Carpenter JW. Exotic Animal Formulary. London: Saunders, 2012.

33. Casadevall A, Pirofski LA. 2001. Host-pathogen interactions: The attributes of virulence. J Infect Dis. 184:337–44.

34. Casella, G., & Berger, R. L. 2002. Statistical inference. Belmont, CA: Duxbury.

35. Castanheira, S., & García-del Portillo, F. 2017. Salmonella Populations inside Host Cells. Frontiers in Cellular and Infection Microbiology, 7.

36. CDC. National Centre for Health Statistics. Life Expectancy. https://www.cdc.gov/nchs/fastats/life-expectancy.htm . [Accessed on 15 October 2017].

37. Center for Disease Control. 2019 Antibiotic Resistance Threats in the United States. Atlanta, GA: U.S. Department of Health and Human Services, CDC; 2019.

38. Chaloner, K., & Verdinelli, I. 1995. Bayesian Experimental Design: A Review. Statistical Science, 10(3), 273-304.

39. Chernoff, Herman. Locally Optimal Designs for Estimating Parameters. Ann. Math. Statist. 24. 1953. no. 4, 586--602.

40. Chisholm RH, Tanaka MM. 2016. The emergence of latent infection in the early evolution of Mycobacterium tuberculosis. Proc Biol Sci. 283:20160499.

41. Cilfone NA, Perry CR, Kirschner DE et al. 2013. Multi-scale modeling predicts a balance of tumor necrosis factor-α and interleukin-10 controls the granuloma environment during Mycobacterium tuberculosis infection. PLoS One. 8:e68680.

42. Ciupe, S. and Heffernan, J. 2017. In-host modeling. Infectious Disease Modelling, 2(2), pp.188-202.

43. Claudi, B., Spröte, P., Chirkova, A., Personnic, N., Zankl, J., Schürmann, N., Schmidt, A., & Bumann, D. 2014. Phenotypic Variation of Salmonella in Host Tissues Delays Eradication by Antimicrobial Chemotherapy. Cell, 158(4), 722–733.

44. Clem, A. 2011. Fundamentals of vaccine immunology. Journal of Global Infectious Diseases, 3(1), 73.

45. Coates, J., Park, B. R., Le, D., Şimşek, E., Chaudhry, W., & Kim, M. 2018. Antibiotic-induced population fluctuations and stochastic clearance of bacteria. eLife, 7, e32976. DOI: 10.7554/eLife.32976

46. Concordet, D., & Nunez, O. G. 2002. A simulated pseudo-maximum likelihood estimator for nonlinear mixed models. Computational Statistics & Data Analysis, 39(2), 187–201.

47. Cooper NG, Julius AA. Bacterial persistence: mathematical modeling and optimal treatment strategy. Proceedings of the 2011 American Control Conference. 2011.

48. Cornforth, D. M., Matthews, A., Brown, S. P., & Raymond, B. 2015. Bacterial Cooperation Causes Systematic Errors in Pathogen Risk Assessment due to the Failure of the Independent Action Hypothesis. PLOS Pathogens, 11(4), e1004775. https://doi.org/10.1371/journal.ppat.1004775

49. Coward C, Restif O, Dybowski R et al. . 2014. The effects of vaccination and immunity on bacterial infection dynamics In Vivo. PLoS Pathog. 10:e1004359.

50. Cronly-Dillon, S. 1972. The relative potencies of heat-killed and acetone-killed vaccines against Salmonella typhimurium in mice. Epidemiology and Infection, 70(4), 597–603.

51. Crum-Cianflone, N. F., & Sullivan, E. 2017. Vaccinations for the HIV-Infected Adult: A Review of the Current Recommendations, Part I. Infectious Diseases and Therapy, 6(3), 303–331.

52. Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., & François, O. 2010. Approximate Bayesian Computation (ABC) in practice. Trends in Ecology & Evolution, 25(7), 410–418.

53. Cui, P., Niu, H., Shi, W., Zhang, S., Zhang, W., & Zhang, Y. 2018. Identification of Genes Involved in Bacteriostatic Antibiotic-Induced Persister Formation. Frontiers in Microbiology, 9.

54. Curtiss, R., III. 2002. Bacterial infectious disease control by vaccine development. Journal of Clinical Investigation, 110(8), 1061–1066.

55. Davies, D. H. 2012. Antigen Discovery for Vaccines Using High-Throughput Proteomic Screening Technologies. In Vaccinology (pp. 150–167). Wiley-Blackwell.

56. Davies, N. G., Flasche, S., Jit, M., & Atkins, K. E. 2019. Within-host dynamics shape antibiotic resistance in commensal bacteria. Nature Ecology & Evolution, 3(3), 440–449

57. Dewachter, L., Fauvart, M., & Michiels, J. 2019. Bacterial Heterogeneity and Antibiotic Survival: Understanding and Combatting Persistence and Heteroresistance. Molecular Cell, 76(2), 255–267.

58. Dhar, N. and McKinney, J. 2007. Microbial phenotypic heterogeneity and antibiotic tolerance. Current Opinion in Microbiology, 10(1), pp.30-38.

59. Diard, M., Bakkeren, E., Hoces, D., Lentsch, V., Arnoldini, M., Böhi, F., Schumann-Moor, K., Adamcik, J., Piccoli, L., Lanzavecchia, A., Stadtmueller, B. M., Donohue, N., van der Woude, M. W., Hockenberry, A., Viollier, P. H., Falquet, L., Wüthrich, D., Bonfiglio, F., Loverdo, C., … Slack, E. 2019. Tailored intestinal IgA responses can set an evolutionary trap for Salmonella Typhimurium. Cold Spring Harbor Laboratory.

60. Diard, M., Sellin, M. E., Dolowschiak, T., Arnoldini, M., Ackermann, M., & Hardt, W.-D. 2014. Antibiotic Treatment Selects for Cooperative Virulence of Salmonella Typhimurium. Current Biology, 24(17), 2000–2005.

61. Doron, S., & Davidson, L. E. 2011. Antimicrobial stewardship. Mayo Clinic proceedings, 86(11), 1113–1123.

62. Drlica, K., Malik, M., Kerns, R. J., & Zhao, X. 2007. Quinolone-Mediated Bacterial Death. Antimicrobial Agents and Chemotherapy, 52(2), 385–392.

63. Du X, King AA, Woods RJ et al. 2017. Evolution-informed forecasting of seasonal influenza A (H3N2). Sci Transl Med. 9:eaan5325.

64. Duijzer, L. E., van Jaarsveld, W., & Dekker, R. 2018. Literature review: The vaccine supply chain. European Journal of Operational Research, 268(1), 174–192.

65. Duneau, D., Ferdy, J. B., Revah, J., Kondolf, H., Ortiz, G. A., Lazzaro, B. P., & Buchon, N. 2017. Stochastic variation in the initial phase of bacterial infection predicts the probability of survival in D. melanogaster. eLife, 6, e28298.

66. Durham, G., & Gallant, A. 2002. [Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes]: Reply. Journal of Business & Economic Statistics, 20(3), 335-338.

67. Dybowski, R., Restif, O., Goupy, A., Maskell, D. J., Mastroeni, P., & Grant, A. J. 2015. Single passage in mouse organs enhances the survival and spread of Salmonella enterica. Journal of The Royal Society Interface, 12(113), 20150702.

68. Dybowski, R., Restif, O., Price, D. J., & Mastroeni, P. 2017. Inferring within-host bottleneck size: A Bayesian approach. Journal of Theoretical Biology, 435, 218–228.

69. Ehrenfeld, E. 1955. On the efficiency of experimental design, Annals of Math. Stat.,26247–255.

70. Elsner, H. L. 1910. THE NEW TREATMENT OF SYPHILIS (EHRLICH-HATA). Journal of the American Medical Association, 55(24), 2052.

71. Endesfelder U. 2019. From single bacterial cell imaging towards in vivo single-molecule biochemistry studies. Essays Biochem. 63:187–96.

72. Erova, T. E., Kirtley, M. L., Fitts, E. C., Ponnusamy, D., Baze, W. B., Andersson, J. A., Cong, Y., Tiner, B. L., Sha, J., & Chopra, A. K. 2016. Protective Immunity Elicited by Oral Immunization of Mice with Salmonella enterica Serovar Typhimurium Braun Lipoprotein (Lpp) and Acetyltransferase (MsbB) Mutants. Frontiers in Cellular and Infection Microbiology, 6.

73. Eykhoff P. System Identification: Parameter and State Estimation. London: Wiley-Interscience, 1974.

74. Fabre F, Montarry J, Coville J et al. 2012. Modelling the evolutionary dynamics of viruses within their hosts: a case study using high-throughput sequencing. PLoS Pathog. 8:e1002654.

75. Falkow, S. 2004. Molecular Koch's postulates applied to bacterial pathogenicity — a personal recollection 15 years later. Nature Reviews Microbiology, 2(1), 67–72.

76. Fantin, B., Ebert, S., Leggett, J., Vogelman, B., & Craig, W. A. 1991. Factors affecting duration of in-vivo postantibiotic effect for aminoglycosides against Gram-negative bacilli. Journal of Antimicrobial Chemotherapy, 27(6), 829–836.

77. Feasey, N. A., Gaskell, K., Wong, V., Msefula, C., Selemani, G., Kumwenda, S., Allain, T. J., Mallewa, J., Kennedy, N., Bennett, A., Nyirongo, J. O., Nyondo, P. A., Zulu, M. D., Parkhill, J., Dougan, G., Gordon, M. A., & Heyderman, R. S. 2015. Rapid Emergence of Multidrug Resistant, H58-Lineage Salmonella Typhi in Blantyre, Malawi. PLOS Neglected Tropical Diseases, 9(4), e0003748.

78. Feasey, N., Dougan, G., Kingsley, R., Heyderman, R. and Gordon, M. 2012. Invasive non-typhoidal salmonella disease: an emerging and neglected tropical disease in Africa. The Lancet, 379(9835), pp.2489-2499.

79. Fenwick, N., Griffin, G., & Gauthier, C. 2009. The welfare of animals used in science: how the "Three Rs" ethic guides improvements. The Canadian veterinary journal = La revue veterinaire canadienne, 50(5), 523–530.

80. Ferro BE, van Ingen J, Wattenberg M, van Soolingen D, Mouton JW. 2015. Time-kill kinetics of antibiotics active against rapidly growing mycobacteria. Journal of Antimicrobial Chemotherapy 70:811–817.

81. Fisher R.A. 1922. On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 222(594–604), 309–368.

82. Fleming A. 1929. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzæ. British journal of experimental pathology, 10(3), 226–236.

83. Frenoy A, Bonhoeffer S. Death and population dynamics affect mutation rate estimates and evolvability under stress in bacteria. PLoS Biol. 2018;16:e2005056

84. Fujiwara M, Takada T. 2017. Environmental Stochasticity. Els.1–8.

85. Gao Y, Li H. 2018. Quantifying and comparing bacterial growth dynamics in multiple metagenomic samples. Nat Methods. 15:1041–4.

86. Garon, J. R., & Orenstein, W. A. 2015. Understanding the host–pathogen interaction saves lives: lessons from vaccines and vaccinations. Current Opinion in Immunology, 36, 8–13.

87. Garre, A., Peñalver-Soto, J. L., Esnoz, A., Iguaz, A., Fernandez, P. S., & Egea, J. A. 2019. On the use of in-silico simulations to support experimental design: A case study in microbial inactivation of foods. PLOS ONE, 14(8), e0220683.

88. Gerlini, A., Colomba, L., Furi, L., Braccini, T., Manso, A. S., Pammolli, A., Wang, B., Vivi, A., Tassini, M., van Rooijen, N., Pozzi, G., Ricci, S., Andrew, P. W., Koedel, U., Moxon, E. R., & Oggioni, M. R. 2014. The Role of Host and Microbial Factors in the Pathogenesis of Pneumococcal Bacteraemia Arising from a Single Bacterial Cell Bottleneck. PLoS Pathogens, 10(3), e1004026.

89. Ghooi, R. B., & Thatte, S. M. 1995. Inhibition of cell wall synthesis — is this the mechanism of action of penicillins? Medical Hypotheses, 44(2), 127–131.

90. Gillespie D. 1977. Exact stochastic simulation of coupled chemical reactions. J Phys Chem. 81:2340–61.

91. Gjedde, S. B., & Gjeode, A. (1980). Organ blood flow rates and cardiac output of the BALB/c mouse. Comparative Biochemistry and Physiology Part A: Physiology, 67(4), 671–674. https://doi.org/10.1016/0300-9629(80)90258-3

92. Gjini, E., & Brito, P. H. 2016. Integrating Antimicrobial Therapy with Host Immunity to Fight Drug-Resistant Infections: Classical vs. Adaptive Treatment. PLOS Computational Biology, 12(4), e1004857.

93. Gog JR, Pellis L, Wood JL et al. 2015. Seven challenges in modeling pathogen dynamics within-host and across scales. Epidemics. 10:45–48.

94. Goh, Y. S., & L Armour, K. 2016. Igg Subclasses Targeting the Flagella of Salmonella Enterica Serovar Typhimurium Can Mediate Phagocytosis and Bacterial Killing. Journal of Vaccines & Vaccination, 07(03).

95. Gollan, B., Grabe, G., Michaux, C., & Helaine, S. 2019. Bacterial Persisters and Infection: Past, Present, and Progressing. Annual Review of Microbiology, 73(1), 359–385.

96. Gonzalez, R. J., Weening, E. H., Lane, M. C., & Miller, V. L. 2015. Comparison of Models for Bubonic Plague Reveals Unique Pathogen Adaptations to the Dermis. Infection and Immunity, 83(7), 2855–2861.

97. Gonzalez-Escobedo, G., Marshall, J. and Gunn, J. 2010. Chronic and acute infection of the gall bladder by Salmonella Typhi: understanding the carrier state. Nature Reviews Microbiology, 9(1), pp.9-14.

98. Gordon, M. 2008. Salmonella infections in immunocompromised adults. Journal of Infection, 56(6), pp.413-422.

99. Gourieroux, C., Monfort, A., & Renault, E. 1993. Indirect inference. Journal of Applied Econometrics, 8(S1), S85–S118.

100. Grant, A. J., Restif, O., McKinley, T. J., Sheppard, M., Maskell, D. J., & Mastroeni, P. 2008. Modelling within-Host Spatiotemporal Dynamics of Invasive Bacterial Disease. PLoS Biology, 6(4), e74.

101. Griffin, A. J., Li, L.-X., Voedisch, S., Pabst, O., & McSorley, S. J. 2011. Dissemination of Persistent Intestinal Bacteria via the Mesenteric Lymph Nodes Causes Typhoid Relapse. Infection and Immunity, 79(4), 1479–1488.

102. Grimm, V. 2005. Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. Science, 310(5750), 987–991.

103. Gulig PA, Doyle TJ. 1993. The Salmonella typhimurium virulence plasmid increases the growth rate of salmonellae in mice. Infect Immun. 61:504–11.

104. Gunn, J., Marshall, J., Baker, S., Dongol, S., Charles, R. and Ryan, E. 2014. Salmonella chronic carriage: epidemiology, diagnosis, and gallbladder persistence. Trends in Microbiology, 22(11), pp.648-655.

105. Guo, N., Duan, H., Kachko, A., Krause, B. W., Major, M. E., & Krause, P. R. 2015. Reverse Engineering of Vaccine Antigens Using High Throughput Sequencing-enhanced mRNA Display. EBioMedicine, 2(8), 859–867.

106. Guzman, C. A., Borsutzky, S., Griot-Wenk, M., Metcalfe, I. C., Pearman, J., Collioud, A., Favre, D., & Dietrich, G. 2006. Vaccines against typhoid fever. Vaccine, 24(18), 3804–3811.

107. Hadjichrysanthou C, Cauet E, Lawrence E, Vegvari C, de Wolf F, Anderson RM. 2016. Understanding the within-host dynamics of influenza A virus: from theory to clinical implications. J R Soc Interface 13(119):20160289.

108. Handel A, Longini IM, Antia R. Towards a quantitative understanding of the within-host dynamics of influenza A infections. J R Soc Interface. 2009;7:35–47.

109. Handel, A., Li, Y., McKay, B., Pawelek, K. A., Zarnitsyna, V., & Antia, R. 2018. Exploring the impact of inoculum dose on host immunity and morbidity to inform model-based vaccine design. PLOS Computational Biology, 14(10), e1006505.

110. Harrington, K.A., and Hormaeche, C.E. 1986. Expression of the innate resistance gene Ity in mouse Kupffer cells infected with Salmonella typhimurium in vitro. Microb Pathog 1: 269–274.

111. Harrison, J. A., Villarreal - Ramos, B., Mastroeni, P., Demarco De Hormaeche, R., & Hormaeche, C. E. 1997. Correlates of protection induced by live Aro − Salmonella typhimurium vaccines in the murine typhoid model. Immunology, 90(4), 618–625.

112. Hartig, F., Calabrese, J. M., Reineking, B., Wiegand, T., & Huth, A. 2011. Statistical inference for stochastic simulation models - theory and application. Ecology Letters, 14(8), 816–827.

113. Haugan, M. S., Charbon, G., Frimodt-Møller, N., & Løbner-Olesen, A. 2018. Chromosome replication as a measure of bacterial growth rate during Escherichia coli infection in the mouse peritonitis model. Scientific Reports, 8(1).

114. Haugan, M. S., Løbner-Olesen, A., & Frimodt-Møller, N. 2018. Comparative Activity of Ceftriaxone, Ciprofloxacin, and Gentamicin as a Function of Bacterial Growth Rate Probed by Escherichia coli Chromosome Replication in the Mouse Peritonitis Model. Antimicrobial Agents and Chemotherapy, 63(2), e02133-18

115. Hawk CT, Leary SL, Morris TH et al. Formulary for Laboratory Animals. Oxford: Blackwell, 2005.

116. Heffernan JM, Keeling MJ. 2008. An in-host model of acute infection: measles as a case study. Theor Popul Biol. 73:134–47.

117. Heffernan JM, Keeling MJ. 2009. Implications of vaccination and waning immunity. Proc Biol Sci. 276:2071–80.

118. Helaine S, Thompson JA, Watson KG et al. 2010. Dynamics of intracellular bacterial replication at the single cell level. Proc Natl Acad Sci. 107:3746–51.

119. Helaine, S., & Holden, D. W. 2013. Heterogeneity of intracellular replication of bacterial pathogens. Current Opinion in Microbiology, 16(2), 184–191.

120. Helaine, S., Cheverton, A. M., Watson, K. G., Faure, L. M., Matthews, S. A., & Holden, D. W. 2014. Internalization of Salmonella by Macrophages Induces Formation of Nonreplicating Persisters. Science, 343(6167), 204–208.

121. Higginson, E. E., Simon, R., & Tennant, S. M. 2016. Animal Models for Salmonellosis: Applications in Vaccine Research. Clinical and Vaccine Immunology, 23(9), 746–756.

122. Hofer, U. The cost of antimicrobial resistance. 2018. Nature Reviews Microbiology, 17(1), 3–3.

123. Hoiseth, S., Stocker, B. 1981. Aromatic-dependent Salmonella typhimurium are non-virulent and effective as live vaccines. Nature 291, 238–239

124. Hormaeche CE. 1980. The in vivo division and death rates of Salmonella typhimurium in the spleens of naturally resistant and susceptible mice measured by the superinfecting phage technique of Meynell. Immunology. 41:973–9.

125. Hur KY, Moon JY, Kim SH et al. 2013. Model-based simulation and prediction of an antiviral strategy against influenza A infection. PLoS One. 8:e68235.

126. Itoh, H., Tokumoto, K., Kaji, T. et al. 2019. Development of a high-throughput strategy for discovery of potent analogues of antibiotic lysocin E. Nat Commun 10, 2992.

127. Izhar, M., DeSilva, L., Joysey, H., & Hormaeche, C. 1990. Moderate immunodeficiency does not increase susceptibility to Salmonella typhimurium aroA live vaccines in mice. Infection And Immunity, 58(7), 2258-2261. doi: 10.1128/iai.58.7.2258-2261.1990

128. Johansson, C., & Wick, M. J. 2004. Liver Dendritic Cells Present Bacterial Antigens and Produce Cytokines upon Salmonella Encounter. The Journal of Immunology, 172(4), 2496–2503.

129. Kaiser, P., Regoes, R. R., Dolowschiak, T., Wotzka, S. Y., Lengefeld, J., Slack, E., Grant, A. J., Ackermann, M., & Hardt, W.-D. 2014. Cecum Lymph Node Dendritic Cells Harbor Slow-Growing Bacteria Phenotypically Tolerant to Antibiotic Treatment. PLoS Biology, 12(2), e1001793.

130. Kaiser, P., Slack, E., Grant, A. J., Hardt, W.-D., & Regoes, R. R. 2013. Lymph Node Colonization Dynamics after Oral Salmonella Typhimurium Infection in Mice. PLoS Pathogens, 9(9), e1003532.

131. Kaldalu, N., Hauryliuk, V., & Tenson, T. 2016. Persisters-as elusive as ever. Applied microbiology and biotechnology, 100(15), 6545–6553.

132. Kanvatirth, P., Rossi, O., Restif, O., Blacklaws, B. A., Tonks, P., Grant, A. J., & Mastroeni, P. 2020. Dual role of splenic mononuclear and polymorphonuclear cells in the outcome of ciprofloxacin treatment of Salmonella enterica infections. Journal of Antimicrobial Chemotherapy, 75(10), 2914–2918.

133. Kariuki, S., Gordon, M. A., Feasey, N., & Parry, C. M. 2015. Antimicrobial resistance and management of invasive Salmonella disease. Vaccine, 33, C21–C29.

134. Khan, M. I., Franco-Paredes, C., Sahastrabuddhe, S., Ochiai, R. L., Mogasale, V., & Gessner, B. D. 2017. Barriers to typhoid fever vaccine access in endemic countries. Research and Reports in Tropical Medicine, Volume 8, 37–44.

135. Kim, J.-S., & Wood, T. K. 2016. Persistent Persister Misperceptions. Frontiers in Microbiology, 7.

136. Kirschner D, Pienaar E, Marino S et al. 2017. A review of computational and mathematical modeling contributions to our understanding of Mycobacterium tuberculosis within-host infection and treatment. Current Opinion in Systems Biology. 3:170–85.

137. Kohanski, M. A., Dwyer, D. J., & Collins, J. J. 2010. How antibiotics kill bacteria: from targets to networks. Nature reviews. Microbiology, 8(6), 423–435.

138. Kokko H, Ebenhard T. 1996. Measuring the Strength of Demographic Stochasticity. J Theor Biol. 183:169–78.

139. Kolter, R., Siegele, D. A., & Tormo, A. 1993. The stationary phase of the bacterial life cycle. Annu. Rev. Microbiol. 47, 855–874.

140. Korem T, Zeevi D, Suez J et al. 2015. Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. Science (New York, NY). 349:1101–6.

141. Kuha, J. 2004. AIC and BIC. Sociological Methods & Research, 33(2), 188–229.

142. Lam, L. H., & Monack, D. M. 2014. Intraspecies Competition for Niches in the Distal Gut Dictate Transmission during Persistent Salmonella Infection. PLoS Pathogens, 10(12), e1004527. DOI: 10.1371/journal.ppat.1004527

143. Landman, D. 1997. Management of infections due to resistant enterococci: a review of therapeutic options. Journal of Antimicrobial Chemotherapy, 40(2), 161–170.

144. Lee, A. J., Wang, S., Meredith, H. R., Zhuang, B., Dai, Z., & You, L. 2018. Robust, linear correlations between growth rates and β-lactam–mediated lysis rates. Proceedings of the National Academy of Sciences, 115(16), 4069–4074.

145. Levin BR, Antia R. 2001. Why we dont get sick: the within-host population dynamics of bacterial infections. Science. 292:1112–5.

146. Levine, M. M., & Farag, T. H. 2011. Invasive Salmonella Infections and HIV in Northern Tanzania. Clinical Infectious Diseases, 52(3), 349–351.

147. Levine, M. M., & Sztein, M. B. 2004. Vaccine development strategies for improving immunization: the role of modern immunology. Nature Immunology, 5(5), 460–464.

148. Levine, M., Black, R. and Lanata, C. 1982. Precise Estimation of the Numbers of Chronic Carriers of Salmonella typhi in Santiago, Chile, an Endemic Area. The Journal of Infectious Diseases, 146(6), pp.724-726.

149. Levin-Reisman, I., Brauner, A., Ronin, I., & Balaban, N. Q. 2019. Epistasis between antibiotic tolerance, persistence, and resistance mutations. Proceedings of the National Academy of Sciences, 116(29), 14734–14739.

150. Lewin, C. S., Morrissey, I., & Smith, J. T. 1991. The mode of action of quinolones: The paradox in activity of low and high concentrations and activity in the anaerobic environment. European Journal of Clinical Microbiology & Infectious Diseases, 10(4), 240–248.

151. Lewis, K. 2010. Persister Cells. Annual Review of Microbiology, 64(1), 357–372.

152. Liepe, J., Filippi, S., Komorowski, M., & Stumpf, M. P. H. 2013. Maximizing the Information Content of Experiments in Systems Biology. PLoS Computational Biology, 9(1), e1002888.

153. Lim, C. H., Voedisch, S., Wahl, B., Rouf, S. F., Geffers, R., Rhen, M., & Pabst, O. 2014. Independent Bottlenecks Characterize Colonization of Systemic Compartments and Gut Lymphoid Tissue by Salmonella. PLoS Pathogens, 10(7), e1004270.

154. Lipsitch, M., & Siber, G. R. 2016. How Can Vaccines Contribute to Solving the Antimicrobial Resistance Problem? MBio, 7(3).

155. Lowe, D. E., Ernst, S. M. C., Zito, C., Ya, J., & Glomski, I. J. 2013. Bacillus anthracis Has Two Independent Bottlenecks That Are Dependent on the Portal of Entry in an Intranasal Model of Inhalational Infection. Infection and Immunity, 81(12), 4408–4420.

156. MacLennan, C. A., Martin, L. B., & Micoli, F. 2014. Vaccines against invasiveSalmonelladisease. Human Vaccines & Immunotherapeutics, 10(6), 1478–1493.

157. Maier, L., Diard, M., Sellin, M. E., Chouffane, E.-S., Trautwein-Weidner, K., Periaswamy, B., Slack, E., Dolowschiak, T., Stecher, B., Loverdo, C., Regoes, R. R., & Hardt, W.-D. 2014. Granulocytes Impose a Tight Bottleneck upon the Gut Luminal Pathogen Population during Salmonella Typhimurium Colitis. PLoS Pathogens, 10(12), e1004557.

158. Marin, J.-M., Pudlo, P., Robert, C. P., & Ryder, R. J. 2011. Approximate Bayesian computational methods. Statistics and Computing, 22(6), 1167–1180.

159. Marino S, Kirschner DE. 2004. The human immune response to Mycobacterium tuberculosis in lung and lymph node. J Theor Biol. 227:463–86.

160. Marino S, Pawar S, Fuller CL et al. 2004. Dendritic cell trafficking and antigen presentation in the human immune response to Mycobacterium tuberculosis. J Immunol. 173:494–506.

161. Markov A.A.. 1906. Extension of the law of large numbers to dependent quantities (in Russian). Izvestiia Fiz.-Matem. Obsch. Kazan Univ., (2nd Ser.), 15(1906)

162. Marks, F., von Kalckreuth, V., Aaby, P., Adu-Sarkodie, Y., El Tayeb, M. A., Ali, M., Aseffa, A., Baker, S., Biggs, H. M., Bjerregaard-Andersen, M., Breiman, R. F., Campbell, J. I., Cosmas, L., Crump, J. A., Espinoza, L. M. C., Deerin, J. F., Dekker, D. M., Fields, B. S., Gasmelseed, N., … Wierzba, T. F. 2017. Incidence of invasive salmonella disease in sub-Saharan Africa: a multicentre population-based surveillance study. The Lancet Global Health, 5(3), e310–e323.

163. Martin, C. J., Cadena, A. M., Leung, V. W., Lin, P. L., Maiello, P., Hicks, N., Chase, M. R., Flynn, J. L., & Fortune, S. M. 2017. Digitally Barcoding Mycobacterium tuberculosis Reveals In Vivo Infection Dynamics in the Macaque Model of Tuberculosis. MBio, 8(3).

164. Mastroeni, P. 2002. Immunity to Systemic Salmonella Infections. Current Molecular Medicine, 2(4), 393–406.

165. Mastroeni, P., Villarreal-Ramos, B., & Hormaeche, C. E. 1992. Role of T cells, TNFα and IFNγ in recall of immunity to oral challenge with virulent salmonellae in mice vaccinated with live attenuated aro− salmonella vaccines. Microbial Pathogenesis, 13(6), 477–491.

166. Mastroeni, P., Villarreal-Ramos, B., & Hormaeche, C. E. 1993. Adoptive transfer of immunity to oral challenge with virulent salmonellae in innately susceptible BALB/c mice requires both immune serum and T cells. Infection and Immunity, 61(9), 3981–3984.

167. Maw J, Meynell GG. 1968. The true division and death rates of Salmonella typhimurium in the mouse spleen determined with superinfecting phage P22. Br J Exp Pathol. 49:597–613.

168. McCall IC, Shah N, Govindan A, Baquero F, Levin BR. 2019. Antibiotic killing of diversely generated populations of nonreplicating bacteria. Antimicrob Agents Chemother 63:e02360-18.

169. McVicker G, Prajsnar TK, Williams A et al. 2014. Clonal expansion during Staphylococcus aureus infection dynamics reveals the effect of antibiotic intervention. PLoS Pathog. 10:e1003959.

170. Mdluli, T., Buzzard, G. T., & Rundell, A. E. 2015. Efficient Optimization of Stimuli for Model-Based Design of Experiments to Resolve Dynamical Uncertainty. PLOS Computational Biology, 11(9), e1004488.

171. Melton-Witt, J. A., Rafelski, S. M., Portnoy, D. A., & Bakardjiev, A. I. 2011. Oral Infection with Signature-Tagged Listeria monocytogenes Reveals Organ-Specific Growth and Dissemination Routes in Guinea Pigs. Infection and Immunity, 80(2), 720–732.

172. Meredith HR, Lopatkin AJ, Anderson DJ et al. 2015. Bacterial temporal dynamics enable optimal design of antibiotic treatment. PLoS Comput Biol. 11:e1004201.

173. Meylan S, Andrews IW, Collins JJ. Targeting antibiotic tolerance, pathogen by pathogen. Cell. 2018;172:1228–38.

174. Meynell GG. 1959. Use of superinfecting phage for estimating the division rate of lysogenic bacteria in infected animals. J Gen Microbiol.21:421–37.

175. Millet YA, Alvarez D, Ringgaard S et al. 2014. Insights into Vibrio cholerae intestinal colonization from monitoring fluorescently labeled bacteria. PLoS Pathog. 10:e1004405.

176. Mittrücker, H.-W., & Kaufmann, S. H. E. 2000. Immune response to infection with Salmonella typhimurium in mice. Journal of Leukocyte Biology, 67(4), 457–463. https://doi.org/10.1002/jlb.67.4.457

177. Moffat, H., Hainy, M., Papanikolaou, N. E., & Drovandi, C. 2020. Sequential experimental design for predator–prey functional response experiments. Journal of The Royal Society Interface, 17(166), 20200156.

178. Monack, D. M., Bouley, D. M., & Falkow, S. 2004. Salmonella typhimurium Persists within Macrophages in the Mesenteric Lymph Nodes of Chronically Infected Nramp1+/+ Mice and Can Be Reactivated by IFNγ Neutralization. Journal of Experimental Medicine, 199(2), 231–241.

179. Moor, K., Diard, M., Sellin, M. E., Felmy, B., Wotzka, S. Y., Toska, A., Bakkeren, E., Arnoldini, M., Bansept, F., Co, A. D., Völler, T., Minola, A., Fernandez-Rodriguez, B., Agatic, G., Barbieri, S., Piccoli, L., Casiraghi, C., Corti, D., Lanzavecchia, A., … Slack, E. 2017. High-avidity IgA protects the intestine by enchaining growing bacteria. Nature, 544(7651), 498–502.

180. Motta S, Pappalardo F. Mathematical modeling of biological systems. Brief Bioinform. 2012;14:411–22.

181. Moxon ER, Murphy PA. 1978. Haemophilus influenzae bacteremia and meningitis resulting from survival of a single organism. Proc Natl Acad Sci. 75:1534–6.

182. Munguia J, Nizet V. 2017. Pharmacological targeting of the host–pathogen interaction: alternatives to classical antibiotics to combat drug-resistant superbugs. Trends Pharmacol Sci. 38:473–88.

183. Myhrvold C, Kotula JW, Hicks WM et al. 2015. A distributed cell division counter reveals growth dynamics in the gut microbiota. Nat Commun. 6:10039.

184. Nelder, J. A., & Mead, R. 1965. A Simplex Method for Function Minimization. The Computer Journal, 7(4), 308–313.

185. Nelder, J. A., & Mead, R. 1965. A Simplex Method for Function Minimization. The Computer Journal, 7(4), 308–313.

186. Nelson, M. L., Dinardo, A., Hochberg, J., & Armelagos, G. J. 2010. Brief communication: Mass spectroscopic characterization of tetracycline in the skeletal remains of an ancient population from Sudanese Nubia 350-550 CE. American Journal of Physical Anthropology, 143(1), 151–154.

187. Nielsen EI, Cars O, Friberg LE. 2011. Predicting in vitro antibacterial efficacy across experimental designs with a semimechanistic pharmacokinetic-pharmacodynamic model. Antimicrobial Agents and Chemotherapy 55:1571–1579.

188. Nilsson, O. R., Kari, L., & Steele-Mortimer, O. 2019. Foodborne infection of mice with Salmonella Typhimurium. PLOS ONE, 14(8), e0215190.

189. Nowak MA, Bangham CRM. 1996. Population dynamics of immune responses to persistent viruses. Science. 272:74–79.

190. O'Malley MA. 2007. Exploratory experimentation and scientific practice: metagenomics and the proteorhodopsin case. Hist Philos Life Sci 29(3):337–360.

191. Ocampo, P. S., Lázár, V., Papp, B., Arnoldini, M., Abel zur Wiesch, P., Busa-Fekete, R., … Bonhoeffer, S. 2014. Antagonism between Bacteriostatic and Bactericidal Antibiotics Is Prevalent. Antimicrobial Agents and Chemotherapy, 58(8), 4573–4582.

192. Ochiai, R., Wang, X., von Seidlein, L., Yang, J., Bhutta, Z., Bhattacharya, S., Agtini, M., Deen, J., Wain, J., Kim, D., Ali, M., Acosta, C., Jodar, L. and Clemens, J. 2005. Salmonella Paratyphi A Rates, Asia. Emerging Infectious Diseases, 11(11), pp.1764-1766.

193. Orwa, T. O., Mbogo, R. W., & Luboobi, L. S. (2019). Multiple-Strain Malaria Infection and Its Impacts on Plasmodium falciparum Resistance to Antimalarial Therapy: A Mathematical Modelling Perspective. Computational and Mathematical Methods in Medicine, 2019, 1–26.

194. Oyston P, Robinson K. 2012. The current challenges for vaccine development. J Med Microbiol. 61(Pt_7):889–94.

195. Pankey, G. A., & Sabath, L. D. 2004. Clinical Relevance of Bacteriostatic versus Bactericidal Mechanisms of Action in the Treatment of Gram - Positive Bacterial Infections. Clinical Infectious Diseases, 38(6), 864–870.

196. Papamakarios, G., C, D., & Murray, I. 2019. Sequential Neural Likelihood: Fast Likelihood-free Inference with Autoregressive Flows. In Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS) 2019 (Vol. 89, pp. 837-848). (Proceedings of Machine Learning Research; Vol. 89). PMLR.

197. Papamakarios, G., Sterratt, D. C., and Murray, I. 2018. Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows. arXiv preprint arXiv:1805.07226.

198. Parry, C., Hien, T., Dougan, G., White, N. and Farrar, J. 2002. Typhoid Fever. New England Journal of Medicine, 347(22), pp.1770-1782.

199. Patel, K. R., Roberts, J. T., & Barb, A. W. 2019. Multiple Variables at the Leukocyte Cell Surface Impact Fc γ Receptor-Dependent Mechanisms. Frontiers in Immunology, 10.

200. Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R. P., Moret, B. M. E., & Stamatakis, A. 2010. How Many Bootstrap Replicates Are Necessary? Journal of Computational Biology, 17(3), 337–354. DOI: 10.1089/cmb.2009.0179

201. Pearson JE, Krapivsky P, Perelson AS. 2011. Stochastic theory of early viral infection: continuous versus burst production of virions. PLoS Comput Biol. 7:e1001058.

202. Penas, D. R., González, P., Egea, J. A., Doallo, R., & Banga, J. R. 2017. Parameter estimation in large-scale systems biology models: a parallel and self-adaptive cooperative strategy. BMC Bioinformatics, 18(1).

203. Pienaar E, Dartois V, Linderman JJ et al. 2015. In silico evaluation and exploration of antibiotic tuberculosis treatment regimens. BMC Syst Biol. 9:79.

204. Pienaar E, Matern WM, Linderman JJ et al. 2016.Multiscale model of Mycobacterium tuberculosis infection maps metabolite and gene perturbations to granuloma sterilization predictions. Infect Immun. 84:1650–69.

205. Pirofski, L. A., & Casadevall, A. 1998. Use of licensed vaccines for active immunization of the immunocompromised host. Clinical microbiology reviews, 11(1), 1–26.

206. Plant, J., & Glynn, A. A. 1976. Genetics of Resistance to Infection with Salmonella typhimurium in Mice. Journal of Infectious Diseases, 133(1), 72–78.

207. Plaut R, Kelly V, Lee G et al. 2012. Dissemination bottleneck in a murine model of inhalational anthrax. Infect Immun. 80:3189–93.

208. Portugal R, Coelho J, Hoper D et al. 2014. Related strains of African swine fever virus with different virulence: genome comparison and analysis. J Gen Virol. 96(Pt_2):408–19.

209. Prajsnar TK, Hamilton R, Garcia-Lara J et al. 2012.A privileged intraphagocyte niche is responsible for disseminated infection of Staphylococcus aureus in a zebrafish model. Cell Microbiol. 14:1600–19.

210. Price, D. J. & Restif, O. 2017. SPEEDI.R - Simulation Package for Efficient Experimental Design and Inference in R. https://github.com/orestif/SPEEDI.R

211. Price, D. J., Bean, N. G., Ross, J. V., & Tuke, J. 2018. Designing group dose-response studies in the presence of transmission. Mathematical Biosciences, 304, 62–78.

212. Price, D. J., Breuzé, A., Dybowski, R., Mastroeni, P., & Restif, O. 2017. An efficient moments-based inference method for within-host bacterial infection dynamics. PLOS Computational Biology, 13(11), e1005841.

213. Pronker ES, Weenen TC, Commandeur H et al. 2013. Risk in vaccine research and development quantified. PLoS One. 8:e57755.

214. Read, M. N., Alden, K., Timmis, J., & Andrews, P. S. 2018. Strategies for calibrating models of biology. Briefings in Bioinformatics.

215. Rego R, Bestor A, Štefka J et al. 2014. Population bottlenecks during the infectious cycle of the lyme disease spirochete Borrelia burgdorferi. PLoS One. 9:e101009.

216. Rego, R. O. M., Bestor, A., Štefka, J., & Rosa, P. A. 2014. Population Bottlenecks during the Infectious Cycle of the Lyme Disease Spirochete Borrelia burgdorferi. PLoS ONE, 9(6), e101009.

217. Regoes RR, Wiuff C, Zappala RM, Garner KN, Baquero F, Levin BR. 2004. Pharmacodynamic functions: A multiparameter approach to the design of antibiotic treatment regimens. Antimicrobial Agents and Chemotherapy 48:3670–3676.

218. Restif, O., Goh, Y. S., Palayret, M., Grant, A. J., McKinley, T. J., Clark, M. R., & Mastroeni, P. 2013. Quantification of the effects of antibodies on the extra- and intracellular dynamics of Salmonella enterica. Journal of The Royal Society Interface, 10(79), 20120866.

219. Restif, O., Thakar, J., & Harvill, E. T. 2009. Analysis with mathematical models provides insights into infectious diseases. Microbe, 4(4), 176-182.

220. Ribet D, Cossart P. 2015. How bacterial pathogens colonize their hosts and invade deeper tissues. Microbes Infect. 17:173–83.

221. Richter-Dahlfors, A., Buchan, A. M. J., & Finlay, B. B. 1997. Murine Salmonellosis Studied by Confocal Microscopy: Salmonella typhimurium Resides Intracellularly Inside Macrophages and Exerts a Cytotoxic Effect on Phagocytes In Vivo. Journal of Experimental Medicine, 186(4), 569–580.

222. Riquelme, S. A., Bueno, S. M., & Kalergis, A. M. 2012. IgG keeps virulent Salmonella from evading dendritic cell uptake. Immunology, 136(3), 291–305.

223. Rong L, Perelson A. 2010. Treatment of Hepatitis C Virus Infection With Interferon and Small Molecule Direct Antivirals: Viral Kinetics and Modeling. Crit Rev Immunol. 30:131–48.

224. Roostalu J, Jõers A, Luidalepp H et al. 2008. Cell division in Escherichia coli cultures monitored at single cell resolution. BMC Microbiol. 8:68.

225. Rosales, C. 2017. Fcγ Receptor Heterogeneity in Leukocyte Functional Responses. Frontiers in Immunology, 8.

226. Rossi, O., Dybowski, R., Maskell, D. J., Grant, A. J., Restif, O., & Mastroeni, P. 2017. Within-host spatiotemporal dynamics of systemic Salmonella infection during and after antimicrobial treatment. Journal of Antimicrobial Chemotherapy, 72(12), 3390–3397.

227. Rossi, O., Vlazaki, M., Kanvatirth, P., Restif, O., & Mastroeni, P. 2020. Within-host spatiotemporal dynamic of systemic salmonellosis: Ways to track infection, reaction to vaccination and antimicrobial treatment. Journal of microbiological methods, 176 106008.

228. Rueckert, C., & Guzmán, C. A. 2012. Vaccines: From Empirical Development to Rational Design. PLoS Pathogens, 8(11), e1003001.

229. Russell WMS, Burch RL. The Principles of Humane Experimental Technique. London, UK: Universities Federation for Animal Welfare; 1959.

230. Sanders, C. C. 1988. Ciprofloxacin: In Vitro Activity, Mechanism of Action, and Resistance. Clinical Infectious Diseases, 10(3), 516–527.

231. Scholar, E. and Pratt, W. 2000. The antimicrobial drugs. Oxford: University Press.

232. Schwartz, D. J., Chen, S. L., Hultgren, S. J., & Seed, P. C. 2011. Population Dynamics and Niche Distribution of Uropathogenic Escherichia coli during Acute and Chronic Urinary Tract Infection. Infection and Immunity, 79(10), 4250–4259.

233. Shaffer M. 1981. Minimum Population Sizes for Species Conservation. Bioscience. 31:131–4.

234. Shehata, T. E., & Marr, A. G. 1971. Effect of nutrient concentration on the growth of Escherichia coli. Journal of bacteriology, 107(1), 210–216.

235. Sheppard M, Webb C, Heath F et al. 2003. Dynamics of bacterial growth and distribution within the liver during Salmonella infection. Cell Microbiol. 5:593–600.

236. Siggins, M. K., Cunningham, A. F., Marshall, J. L., Chamberlain, J. L., Henderson, I. R., & MacLennan, C. A. 2011. Absent Bactericidal Activity of Mouse Serum against Invasive African Non-typhoidal Salmonella Results from Impaired Complement Function but Not a Lack of Antibody. The Journal of Immunology, 186(4), 2365–2371.

237. Silver L. 2011. Challenges of Antibacterial Discovery. Clin Microbiol Rev. 24:71–109.

238. Singh, A. & Hespanha, J. P. 2006. Lognormal moment closures for bio-chemical reactions. In Proc. of the 45th IEEE Conference on Decision and Control, San Diego, CA, 2063–2068.

239. Sinkoe, A., & Hahn, J. 2017. Optimal Experimental Design for Parameter Estimation of an IL-6 Signaling Model. Processes, 5(4), 49.

240. Six, A., Bellier, B., Thomas-Vaslin, V., & Klatzmann, D. 2011. Systems biology in vaccine design. Microbial Biotechnology, 5(2), 295–304.

241. Sneddon, L. U., Halsey, L. G., & Bury, N. R. 2017. Considering aspects of the 3Rs principles within experimental animal biology. The Journal of Experimental Biology, 220(17), 3007–3016.

242. Spall JC. 2005. Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control, volume 65. John Wiley & Sons.

243. Steiert, B., Raue, A., Timmer, J., & Kreutz, C. 2012. Experimental Design for Parameter Estimation of Gene Regulatory Networks. PLoS ONE, 7(7), e40052.

244. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. 2016;69:245–7.

245. Succurro A, Moejes FW, Ebenhöh O. A diverse community to study communities: Integration of experiments and mathematical models to study microbial consortia. J Bacteriol. 2017;199.

246. Sud D, Bigbee C, Flynn J et al. 2006. Contribution of CD8+ T Cells to Control of Mycobacterium tuberculosis Infection. J Immunol. 176:4296–314.

247. Sun, Y., Reid, B., Ferreira, F., Luxardi, G., Ma, L., Lokken, K. L., Zhu, K., Xu, G., Sun, Y., Ryzhuk, V., Guo, B. P., Lebrilla, C. B., Maverakis, E., Mogilner, A., & Zhao, M. 2019. Infection-generated electric field in gut epithelium drives bidirectional migration of macrophages. PLOS Biology, 17(4), e3000044.

248. Tapia, M. D., Tennant, S. M., Bornstein, K., Onwuchekwa, U., Tamboura, B., Maiga, A., Sylla, M. B., Sissoko, S., Kourouma, N., Toure, A., Malle, D., Livio, S., Sow, S. O., & Levine, M. M. 2015. Invasive NontyphoidalSalmonellaInfections Among Children in Mali, 2002–2014: Microbiological and Epidemiologic Features Guide Vaccine Development. Clinical Infectious Diseases, 61(suppl 4), S332–S338.

249. Thakur AK. Model: Mechanistic vs Empirical. In: Rescigno A, Thakur AK(eds). New Trends in Pharmacokinetics. In: NATO ASI Series (Series A: Life Sciences), vol 221 Boston, MA: Springer, 1991.

250. Thakur, A., Mikkelsen, H., & Jungersen, G. 2019. Intracellular Pathogens: Host Immunity and Microbial Persistence Strategies. Journal of Immunology Research, 2019, 1–24.

251. The Center for Disease Dynamics Economics and Policy. (2019). Access Barriers to Antibiotics. Washington, DC: The Center for Disease Dynamics Economics and Policy.

252. Tsolis, R. M., Xavier, M. N., Santos, R. L., & Bäumler, A. J. (2011). How To Become a Top Model: Impact of Animal Experimentation on Human Salmonella Disease Research. Infection and Immunity, 79(5), 1806–1814. https://doi.org/10.1128/iai.01369-10

253. Tuomanen, E., Cozens, R., Tosch, W., Zak, O., & Tomasz, A. 1986. The Rate of Killing of Escherichia coli by -Lactam Antibiotics Is Strictly Proportional to the Rate of Bacterial Growth. Microbiology, 132(5), 1297–1304.

254. Uche IV, MacLennan CA, Saul A. 2017. A Systematic Review of the Incidence, Risk Factors and Case Fatality Rates of Invasive Nontyphoidal Salmonella (iNTS) Disease in Africa (1966 to 2014). PLoS Negl Trop Dis. 11(1):e0005118.

255. Uppington, H., Menager, N., Boross, P., Wood, J., Sheppard, M., Verbeek, S., & Mastroeni, P. 2006. Effect of immune serum and role of individual Fcγ receptors on the intracellular distribution and survival ofSalmonella entericaserovar Typhimurium in murine macrophages. Immunology, 119(2), 147–158.

256. Van den Bergh, B., Fauvart, M., & Michiels, J. 2017. Formation, physiology, ecology, evolution and clinical importance of bacterial persisters. FEMS Microbiology Reviews, 41(3), 219–251.

257. Vanlier, J., Tiemann, C. A., Hilbers, P. A., & van Riel, N. A. 2014. Optimal experiment design for model selection in biochemical networks. BMC Systems Biology, 8(1), 20.

258. Verma S, Bhatt K, Lovey A et al. 2019. Transmission phenotype of Mycobacterium tuberculosis strains is mechanistically linked to induction of distinct pulmonary pathology. PLoS Pathog. 15:e1007613.

259. Vidal, S., Tremblay, M. L., Govoni, G., Gauthier, S., Sebastiani, G., Malo, D., Skamene, E., Olivier, M., Jothy, S., & Gros, P. 1995. The Ity/Lsh/Bcg locus: natural resistance to infection with intracellular parasites is abrogated by disruption of the Nramp1 gene. Journal of Experimental Medicine, 182(3), 655–666.

260. Vlazaki, M., Huber, J., & Restif, O. 2019. Integrating mathematical models with experimental data to investigate the within-host dynamics of bacterial infections. Pathogens and Disease, 77(8).

261. Vlazaki M., Price D.J. & Restif O. 2020 An experimental design tool to optimise inference precision in data-driven mathematical models of bacterial infections in vivo. J. R. Soc. Interface. *In press*.

262. Vlazaki, M., Rossi, O., Price, D. J., McLean, C., Grant, A. J., Mastroeni, P., & Restif, O. 2020. A data-based mathematical modelling study to quantify the effects of ciprofloxacin and ampicillin on the within-host dynamics of Salmonella enterica during treatment and relapse. Journal of The Royal Society Interface.

263. Vosti K. 1989. Serum bactericidal test: past, present, and future use in the management of patients with infections. Curr Clin Top Infect Dis.10:43-55.

264. Walters, M. S., Lane, M. C., Vigil, P. D., Smith, S. N., Walk, S. T., & Mobley, H. L. T. 2012. Kinetics of Uropathogenic Escherichia coli Metapopulation Movement during Urinary Tract Infection. MBio, 3(1).

265. Weill F-X, Domman D, Njamkepo E et al. 2019. Publisher correction: genomic insights into the 2016–2017 cholera epidemic in Yemen. Nature. 566:E14.

266. White, A., Tolman, M., Thames, H. D., Withers, H. R., Mason, K. A., & Transtrum, M. K. 2016. The Limitations of Model-Based Experimental Design and Parameter Estimation in Sloppy Systems. PLOS Computational Biology, 12(12), e1005227.

267. Wiegand, T., Jeltsch, F., Hanski, I., & Grimm, V. 2003. Using pattern-oriented modeling for revealing hidden information: a key for reconciling ecological theory and application. Oikos, 100(2), 209–222.

268. Wilson, G. S. and A. A. Miles. 1964. Topley and Wilson's Principles of Bacteriology and Immunity. FreeWilson and Miles, 1964.

269. Wolkenhauer, O. 2014. Why model?. Frontiers in Physiology, 5.

270. Woods, C., Murdoch, D., Zimmerman, M., Glover, W., Basnyat, B., Wolf, L., Belbase, R. and Reller, L. 2006. Emergence of Salmonella enterica serotype Paratyphi A as a major cause of enteric fever in Kathmandu, Nepal. Transactions of the Royal Society of Tropical Medicine and Hygiene, 100(11), pp.1063-1067.

271. World Health Organization 2018. WHO Position Paper - Weekly epidemiological record. [online] Available at: http://www.who.int/wer/2008/wer8306.pdf?ua=1 (Accessed on 26 Jul. 2018).

272. World Health Organization. 2016. Global Health Estimates: Deaths by Cause, Age, Sex.

273. World Health Organization. World Health Statistics 2019. https://www.who.int/gho/publications/world_health_statistics/2019/EN_WHS_2019_Main.pdf?ua=1 (Accessed on 01 October 2019).

274. Worley, M. J., Nieman, G. S., Geddes, K., & Heffron, F. 2006. Salmonella typhimurium disseminates within its host by manipulating the motility of infected cells. Proceedings of the National Academy of Sciences, 103(47), 17915–17920.

275. Worley, M. J., Nieman, G. S., Geddes, K., & Heffron, F. 2006. Salmonella typhimurium disseminates within its host by manipulating the motility of infected cells. Proceedings of the National Academy of Sciences, 103(47), 17915–17920.

276. Wotzka, S. Y., Kreuzer, M., Maier, L., Arnoldini, M., Nguyen, B. D., Brachmann, A. O., Berthold, D. L., Zünd, M., Hausmann, A., Bakkeren, E., Hoces, D., Gül, E., Beutler, M., Dolowschiak, T., Zimmermann, M., Fuhrer, T., Moor, K., Sauer, U., Typas, A., … Hardt, W.-D. 2019. Escherichia coli limits Salmonella Typhimurium infections after diet shifts and fat-mediated microbiota perturbation in mice. Nature Microbiology, 4(12), 2164–2174.

277. Yan, M., Li, X., Liao, Q., Li, F., Zhang, J. and Kan, B. 2016. The emergence and outbreak of multidrug-resistant typhoid fever in China. Emerging Microbes & Infections, 5(6), pp.e62-e62.

278. Yao, Z., Kahne, D., & Kishony, R. 2012. Distinct Single-Cell Morphological Dynamics under Beta-Lactam Antibiotics. Molecular Cell, 48(5), 705–712.

279. Yoshikawa TT. 2002. Antimicrobial resistance and aging: beginning of the end of the antibiotic era? J Am Geriatr Soc. 50:226–9.

280. Zeiler HJ, Voigt WH. 1987. Efficacy of ciprofloxacin in stationary-phase bacteria in vivo. The American Journal of Medicine. Apr;82(4A):87-90.

281. Zhang X, Zhao H, Vynnycky E et al. 2019. Positively interacting strains that co-circulate within a network structured population induce cycling epidemics of Mycoplasma pneumoniae. Sci Rep. 9:541.

282. Zhang, J. F., Papanikolaou, N. E., Kypraios, T., & Drovandi, C. C. 2018. Optimal experimental design for predator–prey functional response experiments. Journal of The Royal Society Interface, 15(144), 20180186.

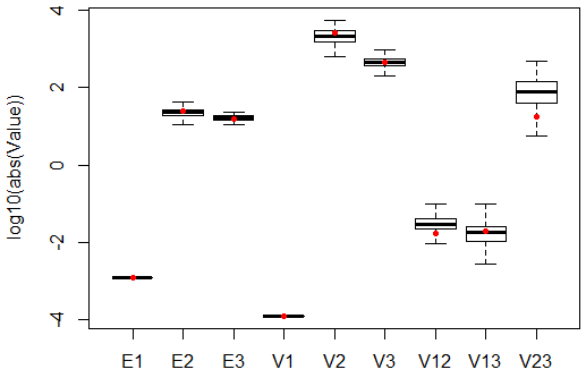
283. Zhang, T., Abel, S., Abel zur Wiesch, P., Sasabe, J., Davis, B. M., Higgins, D. E., & Waldor, M. K. 2017. Deciphering the landscape of host barriers to Listeria monocytogenes infection. Proceedings of the National Academy of Sciences, 114(24), 6334–6339.

9. Appendices

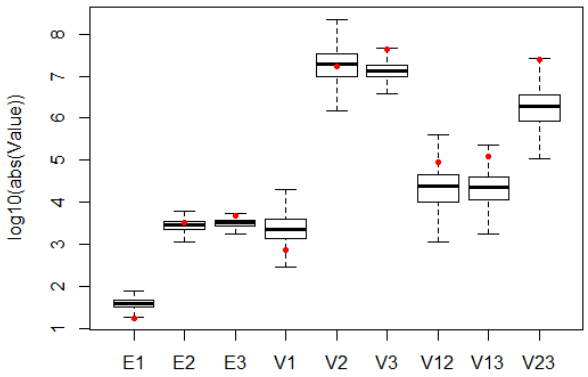
9.1. Model Goodness-of-Fit Supplementary Figures (Chapters 3 and 4)

Chapter 3

Naïve – 24 hours p.i.

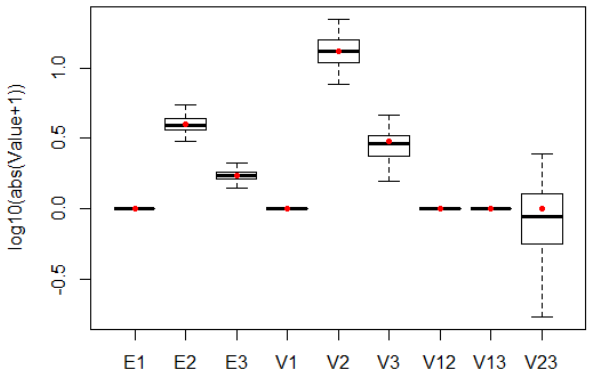


Naïve – 72 hours p.i.

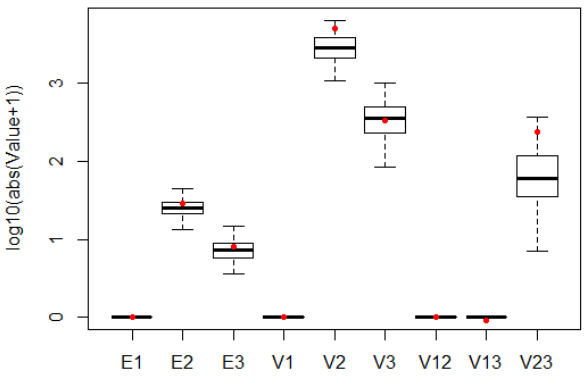


Chapter 4

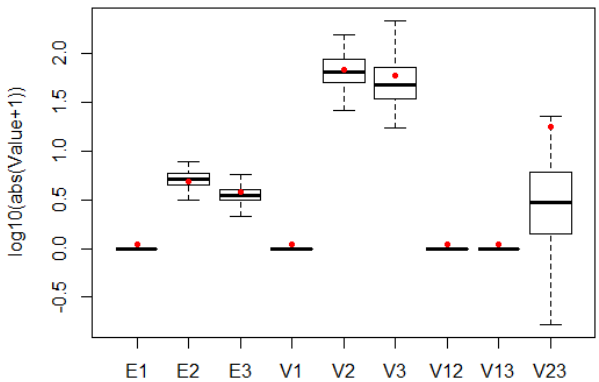
LV group – 6 hours p.i.



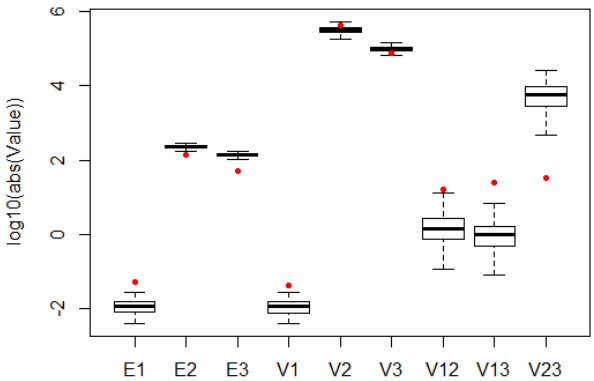
LV group – 48 hours p.i.



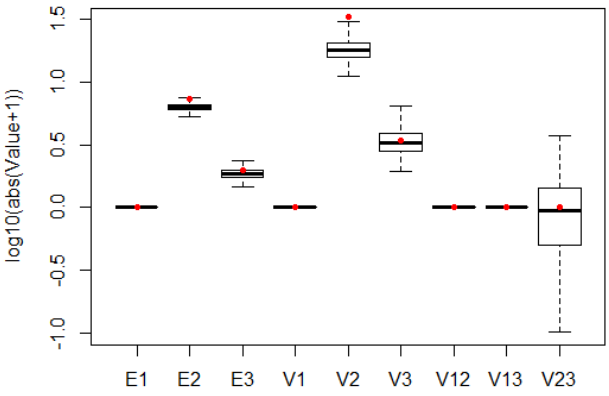
LV group – 24 hours p.i.



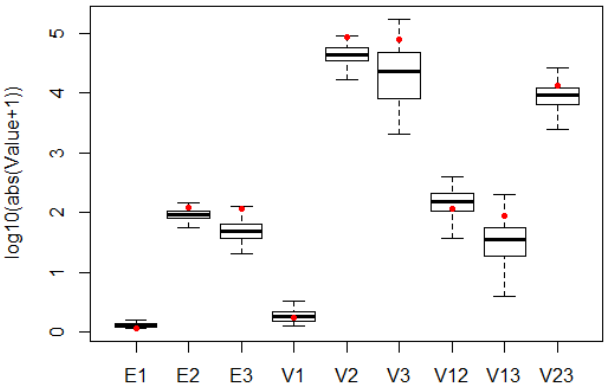
LV group – 72 hours p.i.



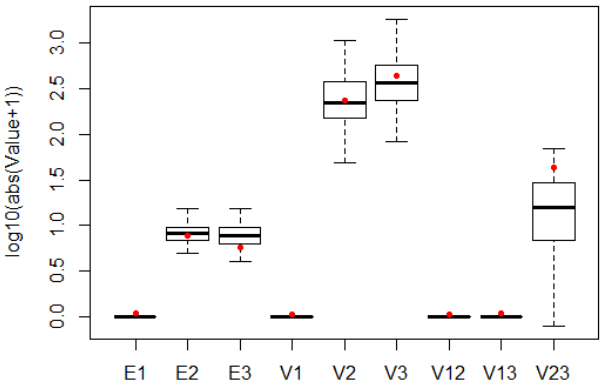
KV group – 6 hours p.i.



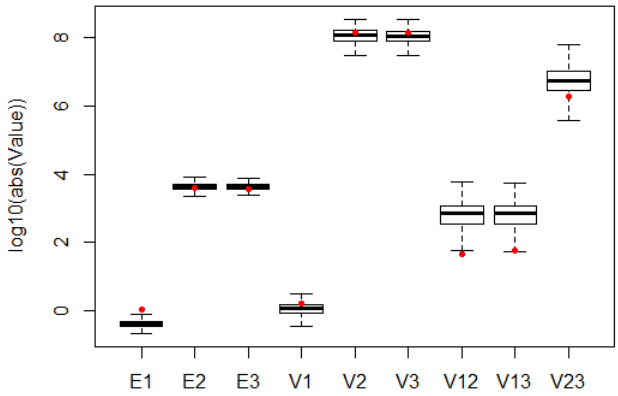
KV group – 48 hours p.i.



KV group – 24 hours p.i.



KV group – 72 hours p.i.



Boxplots of moment values from simulated data (n=500) with red dots corresponding to the experimentally observed moments.

The x-axis refers to the moments and the numbers 1-3 refer to the tissues as follows:
1 – blood
2 – liver
3 – spleen

9.2. Sample Code Implementation

Examples of code implementation for the results obtained in this thesis can be accessed at:

1. For the three-compartment model with a single phenotype (Chapters 3 and 4):
https://github.com/orestif/SPEEDI.R

2. For the four-compartment model with single and dual phenotypes (Chapter 5):
https://github.com/myrtovlazaki/DP_SP_model

3. For the experimental design study (Chapter 6):
https://github.com/myrtovlazaki/OED